
CASE NO. 2023-1663

NONCONFIDENTIAL VERSION

BEFORE THE UNITED STATES
COURT OF APPEALS FOR THE FEDERAL CIRCUIT

STUPP CORPORATION, A DIVISION OF STUPP BROS, INC.,
IPSCO TUBULARS, INC., MAVERICK TUBE CORPORATION,

Plaintiffs,

WELSPUN TUBULAR LLC USA,

Plaintiff-Appellee

v.

UNITED STATES,

Defendant-Appellee,

HYUNDAI STEEL COMPANY,

Defendant,

SEAH STEEL CORP.,

Defendant-Appellant

APPEAL FROM
THE U.S. COURT OF INTERNATIONAL TRADE
IN CASE NOS. 15-CV-00334, 15-CV-00336, AND 15-CV-00337,
JUDGE CLAIRE R. KELLY

NONCONFIDENTIAL BRIEF OF APPELLANT
SEAH STEEL CORPORATION

Submitted by

WINTON & CHAPMAN PLLC
1100 13th Street, N.W., Suite 825
Washington, D.C. 20005
(202) 774-5500

Attorneys for Seah Steel Corporation

July 25, 2023

FORM 9. Certificate of Interest

Form 9 (p. 1)
March 2023

**UNITED STATES COURT OF APPEALS
FOR THE FEDERAL CIRCUIT**

CERTIFICATE OF INTEREST

Case Number 2023-1663

Short Case Caption Stupp Corporation v. U.S.

Filing Party/Entity SeAH Steel Corporation

Instructions:

1. Complete each section of the form and select none or N/A if appropriate.
2. Please enter only one item per box; attach additional pages as needed, and check the box to indicate such pages are attached.
3. In answering Sections 2 and 3, be specific as to which represented entities the answers apply; lack of specificity may result in non-compliance.
4. Please do not duplicate entries within Section 5.
5. Counsel must file an amended Certificate of Interest within seven days after any information on this form changes. Fed. Cir. R. 47.4(c).

I certify the following information and any attached sheets are accurate and complete to the best of my knowledge.

Date: 07/25/2023

Signature: /s/ Jeffrey M. Winton

Name: Jeffrey M. Winton

FORM 9. Certificate of Interest

Form 9 (p. 2)
March 2023

| 1. Represented Entities. Fed. Cir. R. 47.4(a)(1). | 2. Real Party in Interest. Fed. Cir. R. 47.4(a)(2). | 3. Parent Corporations and Stockholders. Fed. Cir. R. 47.4(a)(3). |
|---|--|--|
| Provide the full names of all entities represented by undersigned counsel in this case. | Provide the full names of all real parties in interest for the entities. Do not list the real parties if they are the same as the entities. <input checked="" type="checkbox"/> None/Not Applicable | Provide the full names of all parent corporations for the entities and all publicly held companies that own 10% or more stock in the entities. <input type="checkbox"/> None/Not Applicable |
| SeAH Steel Corporation | N/A | SeAH Steel Holdings Corporation |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

☐ Additional pages attached

FORM 9. Certificate of Interest

Form 9 (p. 3)
March 2023

4. Legal Representatives. List all law firms, partners, and associates that (a) appeared for the entities in the originating court or agency or (b) are expected to appear in this court for the entities. Do not include those who have already entered an appearance in this court. Fed. Cir. R. 47.4(a)(4).

☐ None/Not Applicable

☐ Additional pages attached

| | | |
|---------------------------------|--|--|
| Law Firm: Winton & Chapman PLLC | | |
| | | |
| | | |

5. Related Cases. Other than the originating case(s) for this case, are there related or prior cases that meet the criteria under Fed. Cir. R. 47.5(a)?

☒ Yes (file separate notice; see below)
 ☐ No
 ☐ N/A (amicus/movant)

If yes, concurrently file a separate Notice of Related Case Information that complies with Fed. Cir. R. 47.5(b). **Please do not duplicate information.** This separate Notice must only be filed with the first Certificate of Interest or, subsequently, if information changes during the pendency of the appeal. Fed. Cir. R. 47.5(b).

6. Organizational Victims and Bankruptcy Cases. Provide any information required under Fed. R. App. P. 26.1(b) (organizational victims in criminal cases) and 26.1(c) (bankruptcy case debtors and trustees). Fed. Cir. R. 47.4(a)(6).

☒ None/Not Applicable

☐ Additional pages attached

Table of Contents

| | <u>Page</u> |
|--|-------------|
| STATEMENT OF RELATED CASES | 1 |
| JURISDICTIONAL STATEMENT..... | 1 |
| STATEMENT OF ISSUES..... | 2 |
| STATEMENT OF THE CASE | 3 |
| SUMMARY OF ARGUMENT..... | 8 |
| A. Commerce’s Response to This Court’s Questions Concerning the Use of Cohen’s d When the Data Being Analyzed Does Not Meet the Recognized Requirements for Using that Test | 8 |
| B. Commerce’s Response to This Court’s Observation that, under Cohen’s d as Used by Commerce, an Imperceptible Price Difference May Be Found to Be “Significant”..... | 9 |
| C. The Effect of the DPA on Commerce’s Past Determinations as a Measure of Its Reasonableness | 11 |
| D. Heights of Teenaged Girls or the IQs of Different Types of Students as Universal Yardsticks..... | 13 |
| ARGUMENT | 15 |
| A. Commerce’s Redetermination Fails to Address the Flaws in Commerce’s Use of Cohen’s d that Were Identified by This Court..... | 15 |
| 1. Commerce Has Not Identified Any Academic Texts that Support Its Claim that Cohen’s d May Be Used When the Assumptions Underlying that Test Are Not Satisfied | 15 |

Page

| | | |
|----|---|----|
| 2. | Commerce Has Not Identified Any Logical or Factual Support for Its Claimed Use of Cohen’s d when Professor Cohen’s Assumptions Are Not Satisfied | 22 |
| B. | Commerce’s Use of the Cohen’s d Test Can Classify Imperceptible Differences as Significant, Just as this Court Predicted | 25 |
| 1. | Commerce’s Claim that the “Meaningful Difference” Test Takes Care of the Situations Described by the Court’s Hypothetical Example Is Incorrect When There Is More than One Product Being Analyzed..... | 25 |
| 2. | Commerce’s “Ratio Test” Cannot Give Meaningful Results when the Cohen’s d Test Fails to Properly Identify Significant Price Differences | 32 |
| C. | Commerce’ Attempt to Show that Its Use of the Differential Pricing Analysis Affected the Outcomes in Only a Small Number of Investigations Misrepresents the Actual Impact of Commerce’s Methodology and Is Fundamentally Irrelevant..... | 37 |
| D. | Commerce’s Assertion that Professor Cohen’s Proposed Thresholds Can Be Used as Universal Yardsticks Because They Are Based on Real-World Observations Is Illogical and Contrary to the Evidence | 42 |
| 1. | Professor Cohen Expressly Rejected the Notion that His Rules-of-Thumb Established Universal Yardsticks | 43 |
| 2. | There Is No Reason to Expect SeAH’s U.S Prices to Follow a Similar Distribution to Heights or IQs | 46 |
| 3. | The Evidence Confirms that There Are No Discernible Differences in the Prices for Products that, According to Commerce, Show a “Large” Effect Size under Its Cohen’s d Test..... | 48 |

Page

| | |
|-----------------|----|
| CONCLUSION..... | 54 |
|-----------------|----|

CONFIDENTIAL MATERIAL OMMITTED

Respondent's confidential sales information is omitted from footnote 68 on page 47, lines 7, 8, and 10 through 13 on page 49, and lines 3 through 5 on page 50.

Diagrams representing Respondent's confidential sales information are omitted from pages 50 and 51.

ADDENDUM

1. *Stupp Corp. et al. v. United States*, Slip Op. 23-23 (CIT, Feb. 24, 2023)

Table of Authorities

| | <u>Page</u> |
|---|---------------|
| <u>U.S. COURT DECISIONS</u> | |
| <i>Creswell Trading Co. Inc. v. U.S.</i> , 15 F.3d 1054, 1060–61 (Fed. Cir. 1994) | 19 |
| <i>Stupp Corp. v. United States</i> , 5 F.4th 1341 (Fed. Cir. 2021) | <i>passim</i> |
| <i>Virnetx, Inc. v. Cisco Sys., Inc.</i> , 767 F.3d 1308 (Fed. Cir. 2014) | 15 |
| <u>STATUTES</u> | |
| 19 U.S.C. § 1677f-1(d)(1)(B) | 19 |
| <u>OTHER AUTHORITIES</u> | |
| 1 <i>Oxford Handbook of Quantitative Methods</i> (2013)..... | 21 |
| Cohen, STATISTICAL POWER ANALYSIS FOR THE BEHAVIORAL SCIENCES (2d ed. 1988)..... | <i>passim</i> |
| Ellis, Paul, THE ESSENTIAL GUIDE TO EFFECT SIZES: STATISTICAL POWER, META-ANALYSIS, AND THE INTERPRETATION OF RESEARCH RESULTS (2010)..... | 35, 44 |
| Hedges and Olkin, <i>Overlap Between Treatment and Control Distributions as an Effect Size Measure in Experiments</i> , 21:1 PSYCHOLOGICAL METHODS (2016) | 17 |
| Huberty and Lohman, <i>Group Overlap as a Basis for Effect Size</i> , 60:4 EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT (2000)..... | 17 |
| J. Cohen, <i>A Power Primer</i> , 112:1 PSYCHOLOGICAL BULLETIN (1992) | 44 |

Page

| | |
|--|------------|
| Karris, Steven, MATHEMATICS FOR BUSINESS, SCIENCE AND TECHNOLOGY (3d ed. 2007) | 21, 24 |
| Ricca and Blaine, <i>Notes on a Nonparametric Estimate of Effect Size</i> , 90:1 JOURNAL OF EXPERIMENTAL EDUCATION (2022)..... | 20 |
| Starnes, Yates, and Moore, <i>Statistics through Applications</i> (2005) | 22, 23, 46 |

NONCONFIDENTIAL BRIEF OF APPELLANT
SEAH STEEL CORPORATION

This brief is submitted on behalf of Appellant SeAH Steel Corporation (“SeAH”) in the appeal of the February 24, 2023, decision by the Court of International Trade (the “CIT”) in *Stupp Corporation et al. v. United States*, Consol. Ct. No. 15-00334.

STATEMENT OF RELATED CASES

Counsel for SeAH is unaware of any cases pending in this court that may directly affect or be directly affected by this Court’s decision in this appeal.

JURISDICTIONAL STATEMENT

This appeal arises out of the final determination by the Department of Commerce in the antidumping investigation of *Welded Line Pipe from Korea*. Notice of Commerce’s final determination was published in the *Federal Register* on October 13, 2015. *See* 80 Fed. Reg. 61366 (Oct. 13, 2015).¹ Commerce’s determination was appealed to the Court of International Trade pursuant to 19 U.S.C. §§ 1516a(a)(2)(A)(i)(II) and (a)(2)(B)(i), and the CIT exercised jurisdiction over that appeal by reason of 28 U.S.C. § 1581(c).

¹ An amendment to that determination was published on November 10, 2015. *See* 80 Fed. Reg. 69637 (Nov 10, 2015) (Appx0195-0196).

The CIT initially entered final judgment in this action on March 24, 2020. That CIT decision was overturned by this Court on appeal, and the case was remanded to the CIT for further proceedings in accordance with this Court's decision. The CIT entered final judgment in the remand proceedings on February 24, 2023. SeAH filed notice of its appeal to this Court 26 days after the entry of the CIT's final judgment (on March 22, 2023), making this appeal timely under Federal Circuit Rule 4. This Court has jurisdiction pursuant to 28 U.S.C. § 1295(a)(5). This appeal is from a final order or judgment by the CIT that disposes of all parties' claims.

STATEMENT OF ISSUES

1. Whether it is reasonable for Commerce to use a statistical test in a manner inconsistent with the limitations on the methodology described by the methodology's creator and relevant academic literature, and without any mathematical, logical, or empirical explanation why such a method may properly be used in the manner Commerce proposes?
2. Whether it is reasonable for Commerce to use a methodology that Commerce itself admits may fail to correctly identify whether price differences are "significant," when the statute requires a finding whether price differences are "significant," and when an incorrect finding that price differences for specific sales are "significant" necessarily distorts the count of

the total number of sales with “significant” price differences (for purposes of Commerce’s “Ratio Test”) and, as a consequence, the results of Commerce’s “differential pricing analysis”?

3. Whether the overall number of cases in which the results were affected by Commerce’s use of its “differential pricing analysis” can, in a vacuum, demonstrate the reasonableness of that methodology, when there is no evidence as to how many cases actually satisfied the relevant statutory criteria?

4. Whether rules-of-thumb supported by observations of the heights of teenaged girls and the IQs of different types of students provide a universal yardstick for determining whether observed price differences are “large,” when the prices are not set based on heights or IQs, and the distribution of prices is not the same as the distribution of heights and IQs?

STATEMENT OF THE CASE

The antidumping statute generally requires Commerce to calculate the dumping margins for individual products (which Commerce defines by reference to “control numbers” or “CONNUMs”) by comparing the average price for the U.S. sales to the average price for sales in the comparison market. However, the statute permits Commerce to depart from the usual “average-to-average” (or “A-to-A” comparison) if it finds (1) that “there is a

pattern of {U.S. prices} for comparable merchandise that differ significantly among purchasers, regions, or periods of time,” and (2) that this “pattern” cannot be “taken into account” by the A-to-A comparison. When those statutory requirements are met, Commerce is authorized to compare the prices for individual U.S. sales transactions to the average price for sales in the comparison market. The arithmetic of that “A-to-T” comparison tends to increase dumping margins, in large part because Commerce treats any negative dumping margins on individual sales as if they were zero when calculating the overall margin (in a process known as “zeroing”).²

In order to address this statutory provision, Commerce has adopted something it calls the Differential Pricing Analysis (“DPA”). The DPA consists of three steps:

- (1) First, Commerce groups the sales of each product by customer, region, and time period, compares the average price for each such group to the average price for all other sales of the same product, and then calculates a figure known as Cohen’s d by dividing the difference in average prices by the average standard deviation of

² In response to arguments by another respondent, Commerce stated that “the differences in the calculated weighted-average dumping margins are solely due to zeroing or the denial of offsets for non-dumped U.S. sales.” *See* Final I&D Memorandum at 14.

prices in the group being examined and in the group consisting of all other sales of the product. If the calculated Cohen's d for a given sale is found to be greater than 0.8, Commerce concludes that the sale has "passed" the Cohen's d test.

- (2) Next, Commerce calculates the total value of respondent's U.S. sales that passed the Cohen's d test, and divides that amount by the total value of all of the respondent's U.S. sales. If this "Ratio Test" results in a quotient that is less than 33 percent, Commerce concludes that the A-to-A methodology should be used for comparison. If the quotient is greater than 66 percent, Commerce concludes that the A-to-T methodology should be used for comparison for all U.S. sales. If the quotient is between 33 and 66 percent, Commerce concludes that a "mixed" methodology should be used that applies the A-to-T methodology should be used for comparison, but only for sales that "passed" the Cohen's d test.

- (3) Finally, Commerce compares the dumping margin that would be calculated using the ordinary A-to-A comparison for all sales with the dumping margin that would be calculated using the comparison methodology dictated by the "Ratio Test." If the resulting dumping margin crosses from below the *de minimis* threshold to above the *de*

minimis threshold (or if an above *de minimis* dumping margin increases by at least 25 percent), Commerce concludes that there is a “meaningful difference” in the results, and bases its determination on the results of the comparison dictated by the “Ratio Test.” If no “meaningful difference” is found, Commerce bases its determination on the results of the ordinary A-to-A test. For these purposes, Commerce considers a dumping margin below 2 percent to be *de minimis* in an investigation, and a dumping margin below 0.5 percent to be *de minimis* in a review.

In the investigation of Welded Line Pipe from Korea, the dumping margin for SeAH would have been a *de minimis* 1.97 percent if the A-to-A comparison methodology had been used.³ However, Commerce’s DPA found that 39.72 percent of SeAH’s U.S. sales “passed” the Cohen’s *d* test. Commerce also calculated that the dumping margin for SeAH under the “mixed” methodology” would be 2.53 percent. Because the dumping margin for SeAH crossed from below *de minimis* under the A-to-A methodology to above *de minimis* under the “mixed” methodology, Commerce based its final

³ See Memorandum re Final Determination Margin Calculation for SeAH at 3 (Oct. 5, 2015) (Appx1449).

determination on the dumping margin generated by the “mixed” methodology, and thus made an affirmative final determination for SeAH.

In its appeal of Commerce’s determination, SeAH challenged Commerce’s reliance on this Differential Pricing Analysis to justify a departure from the A-to-A comparison ordinarily required by the statute. In its previous decision in that appeal (hereinafter, “*Stupp III*”), this Court affirmed various aspects of Commerce’s DPA, but rejected Commerce’s reliance on the Cohen’s *d* test.⁴ As a separate matter, this Court’s decision in *Stupp III* also noted that the manner in which Commerce calculated Cohen’s *d* could result in affirmative findings of patterns when there are only insignificant differences in U.S. prices.⁵ The Court remanded the case to allow Commerce an opportunity to address those issues.

Commerce issued its Redetermination on Remand on April 4, 2022.⁶ SeAH timely filed comments challenging that Redetermination before the Court of International Trade (“CIT”). On February 24, 2023, the CIT issued a

⁴ See *Stupp Corp. v. United States*, 5 F.4th 1341 (Fed. Cir. 2021) (hereinafter “*Stupp III*”).

⁵ See *Stupp III*, 5 F.4th at 1359.

⁶ See Commerce’s April 4 Final Results of Redetermination Pursuant to Court Remand (Appx0028-0101) (hereinafter “Redetermination”).

decision affirming Commerce's Redetermination.⁷ SeAH appealed that decision to this Court, and files this brief in support of that appeal.

SUMMARY OF ARGUMENT

*A. Commerce's Response to This Court's Questions
Concerning the Use of Cohen's d When the
Data Being Analyzed Does Not Meet the
Recognized Requirements for Using that Test*

The academic literature demonstrates that the Cohen's d statistic was intended to be used as a measure of "effect size" only when the data being analyzed consists of samples drawn from Normal distributions, with roughly equal variance, and a sufficient number of data-points. In its decision in *Stupp III*, this Court took note of that academic literature, and questioned Commerce's use of the Cohen's d statistic in circumstances in which those requirements are not satisfied.

In response, Commerce's Redetermination set forth a lengthy discussion that purported to show that Professor Cohen's discussion of his d statistic only addressed the use of that statistic in the analysis of samples, and not situations in which the entire population was being considered.⁸ Commerce's Redetermination also argued that the other texts referenced by this Court in

⁷ See *Stupp Corporation v. United States*, 619 F.Supp.3d 1314 (CIT 2023) (hereinafter "*Stupp IV*").

⁸ See Redetermination at 19-20 (Appx0046-0047).

Stupp III again only addressed the use of the Cohen's d to assess the statistical significance of analyses based on sampled data.⁹ According to Commerce, none of those materials *prohibited* the use of the Cohen's d test for the analysis of entire populations.

Tellingly, Commerce failed to identify any texts that *support* its proposed use of Cohen's d when the assumptions described by Professor Cohen are not satisfied. Moreover, Commerce also failed to provide any independent logical, mathematical, or empirical justification for its use of Cohen's d when the usual assumptions are not satisfied. In these circumstances, Commerce's use of Cohen's d , and of the rule-of-thumb proposed by Professor Cohen for determining whether a particular d value was "large," cannot be justified as application of a widely-adopted statistical practice. If Commerce nevertheless wishes to rely on that rule-of-thumb, it must provide some other basis for concluding that those thresholds are reasonable.

B. Commerce's Response to This Court's Observation that, under Cohen's d as Used by Commerce, an Imperceptible Price Difference May Be Found to Be "Significant"

As mentioned, this Court's decision in *Stupp III* also noted that the manner in which Commerce calculated Cohen's d could result in affirmative

⁹ *Id.* at 20-25 (Appx0047-0052).

findings of patterns when there are only insignificant differences in U.S. prices.¹⁰ In response, Commerce’s Redetermination asserted that the separate “meaningful difference” test applied as part of the DPA addressed any such concern — since insignificant differences in U.S. prices would not result in an increase in dumping margins from zero to above Commerce’s 2 percent *de minimis* threshold.

But, Commerce’s assertion is wrong: An incorrect finding of a “significant” price difference for a particular customer, region, or time period for one product may not necessarily result in a dumping margin *for that product*. But it necessarily will affect the number of transactions found to “pass” Commerce’s Cohen’s *d* test. And, because Commerce’s “Ratio Test” depends on the number of transactions found to “pass” the Cohen’s *d* test, an incorrect count of the number of transactions found to “pass” the Cohen’s *d* test may lead to an alternate comparison methodology being used for other sales, for which the change in methodology will inflate the dumping margin.

Consequently, Commerce’s argument that its “meaningful difference” test takes care of the problem identified by the Court is true, if at all, only for a case involving the analysis of a single product. In a case involving multiple

¹⁰ See *Stupp III*, 5 F.4th at 1359.

products, like the Welded Line Pipe investigation at issue in this appeal, Commerce's argument has no validity.

Furthermore, Commerce's assertion that its "meaningful difference" test compensates for errors in its Cohen's *d* test is inconsistent with Commerce's insistence that each element of its "differential pricing analysis" addresses a separate part of the statutory requirements. If Commerce's assertion is correct, then each element of the DPA must stand on its own. The fact that the Cohen's *d* test fails to correctly identify whether a price difference is "significant" means that the test is inconsistent with the statutory requirement, regardless of whether Commerce concludes that a "pattern" of price differences that were incorrectly identified as "significant" could or could not be taken into account using an A-to-A comparison.

C. The Effect of the DPA on Commerce's Past Determinations as a Measure of Its Reasonableness

Commerce asserted, and the CIT seems to have accepted, that an analysis showing that the DPA only changed Commerce's comparison methodology in 21 or 22 percent of cases demonstrated that the DPA was reasonable. The logical error in that argument should be obvious: Unless one knows how often Commerce *should* have changed its methodology, one cannot assess whether Commerce's actual experience is reasonable or not. The statute does not direct Commerce to change its methodology in a "reasonable number of

cases.” Instead, it directs Commerce to change its methodology only when specific factual conditions have been satisfied. If those conditions are satisfied in only 5 percent of cases, then the fact that the DPA resulted in a change in Commerce’s comparison methodology in 21 or 22 percent of cases would confirm that Commerce’s DPA was not reasonable.

Furthermore, Commerce’s assertion that its use of the DPA “only” changed the comparison methodology in 21 or 22 percent of cases is, itself, misleading. The DPA actually has two different standards: For cases in which the dumping margins would have been above *de minimis* under the A-to-A methodology, an increase in dumping margins of 25 percent is needed to justify a change in comparison methodology. By contrast, for cases in which the dumping margins would have been *de minimis* under the A-to-A methodology, any increase in the dumping margin across the *de minimis* threshold would be sufficient to justify a change in comparison methodology.

SeAH’s dumping margin under the A-to-A methodology would have been a below *de minimis* 1.97 percent. Accordingly, all that was required to justify a change in methodology under Commerce’s DPA was an increase in that margin of 0.03 percent. Statistics about how the DPA affected cases in which a 25 percent increase in dumping margins was required to change the comparison methodology are irrelevant to SeAH’s situation.

If one examines cases, like SeAH's, in which the dumping margin would have been *de minimis* under the A-to-A methodology, the results are quite different than Commerce claims. In fact, the DPA has resulted in changing negative determinations to affirmative determinations in more than 50 percent of such cases since it was introduced. To the extent that such figures can be judged in a vacuum, that result hardly seems reasonable.

*D. Heights of Teenaged Girls or
the IQs of Different Types of
Students as Universal Yardsticks*

In its Redetermination, Commerce claimed that its use of Professor Cohen's rules-of-thumb was justified by the fact the thresholds proposed by Professor Cohen were based on actual observations of real-world data — what Professor Cohen described as “grossly perceptible” differences in the heights of teenage girls of different ages and the IQs of different types of students — that are independent of the intended purpose of his *d* test.¹¹

But the fact that Professor Cohen considered a calculated *d* of 0.8 to correspond to “grossly perceptible” differences in heights or IQ does not mean that they provide a universal yardstick for measuring whether differences in other data are grossly perceptible. In fact, Professor Cohen

¹¹ Redetermination at 17-18 (Appx0044-0045).

himself denied that the d statistic could be used as a universal yardstick, or that his proposed thresholds were appropriate for all areas of research in behavioral science, let alone other fields.

Furthermore, there is no logical reason to expect the calculated d for heights and IQs of different populations to measure whether differences in SeAH's average sales prices are "large." After all, SeAH does not measure the height of its customers or make its customers take IQ tests as part of its price negotiations. And, as a mathematical matter, the distribution of SeAH's prices differs materially from the distribution of heights and IQs. Both heights and IQs are Normally distributed and are measured from data that has a large number of data-points. SeAH's sales prices are neither. Consequently, there is no reason to expect a rule-of-thumb based on observations of heights or IQs to provide any meaningful threshold for SeAH's U.S. prices.

ARGUMENT

- A. *Commerce's Redetermination Fails to Address the Flaws in Commerce's Use of Cohen's d that Were Identified by This Court*
1. *Commerce Has Not Identified Any Academic Texts that Support Its Claim that Cohen's d May Be Used When the Assumptions Underlying that Test Are Not Satisfied*

In its decision in *Stupp III*, this Court held that, when an agency purports to apply a mathematical tool, it must apply that tool in a manner consistent with the tool's assumptions. As the Court explained,

... Professor Cohen derived his interpretive cutoffs under certain assumptions. Violating those assumptions can subvert the usefulness of the interpretive cutoffs, transforming what might be a conservative cutoff into a meaningless comparator.¹²

In *Stupp III*, the Court concluded that “the evidence and arguments before us call into question whether Commerce’s application of the Cohen’s d test to the data in this case violated the assumptions of normality, sufficient observation size, and roughly equal variances associated with that test.”¹³

This Court therefore remanded this case to give Commerce an opportunity to explain “whether the limits on the use of the Cohen’s d test prescribed by

¹² See *Stupp III* at 1360, citing *Virnetx, Inc. v. Cisco Sys., Inc.*, 767 F.3d 1308, 1332 (Fed. Cir. 2014).

¹³ *Id.*

Professor Cohen and other authorities were satisfied in this case or whether those limits need not be observed when Commerce uses the Cohen’s *d* test in less-than-fair-value adjudications.”¹⁴ And, the Court explicitly invited Commerce “to clarify its argument that having the entire universe of data rather than a sample makes it permissible to disregard the otherwise-applicable limitations on the use of the Cohen’s *d* test.”¹⁵

In response to this Court’s decision, Commerce issued a letter requiring SeAH to submit on the record copies of all of the texts that were cited in the *Stupp III* decision.¹⁶ SeAH provided the requested copies.¹⁷ While the domestic parties were given an opportunity to “rebut” SeAH’s submission, their only “rebuttal” filed consisted of two excerpts from books by Professor

¹⁴ *Id.*

¹⁵ *Id.*

¹⁶ *See* Letter to All Interested Parties (Oct. 29, 2021) (Appx3715-3718).

¹⁷ SeAH initially submitted the requested documents on November 5, 2012. However, Commerce rejected SeAH’s submission because SeAH had taken the position that documents obtained under copyright could not be placed in their entirety on the public record and, instead, had to be treated as proprietary documents. A revised submission of the requested documents was filed by SeAH on November 12, 2021 (Appx3719-5435).

Cohen and Professor Ellis (which duplicated materials already included in SeAH's submission).¹⁸

In its Redetermination, Commerce went through the various texts cited by the *Stupp III* decision and attempted to explain why none of those texts addressed the use of Cohen's d in the manner in which it was used in Commerce's DPA.¹⁹ We agree that none of the texts discusses anything like Commerce's DPA.²⁰ None of the texts explicitly addresses the use of

¹⁸ See Welspun Tubular's November 19, 2021, Submission (Appx5436-5457).

¹⁹ See Redetermination at 20-25 (Appx0047-0052).

²⁰ We should note, however, that some of Commerce's attempts to distinguish its use of Cohen's d from the use described in the academic texts are unpersuasive. For example, Commerce claims that "overlap" is somehow analytically distinct from "effect size." See Redetermination at 20 (Appx0047). However, the academic literature makes clear that "overlap" is actually a measure of "effect size." See, e.g., Hedges and Olkin, *Overlap Between Treatment and Control Distributions as an Effect Size Measure in Experiments*, 21:1 PSYCHOLOGICAL METHODS 61 (2016) (Appx6381); Huberty and Lohman, *Group Overlap as a Basis for Effect Size*, 60:4 EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT 543 (2000) (Appx6383).

Furthermore, Commerce's claim that it is analyzing the entire population, and not just a sample, is also questionable: Commerce looks at sales only during a particular time period (coinciding with the investigation period) and not all sales by the exporter over time. And, even within the universe of sales during the period, Commerce does not compare the entire population, but instead analyzes samples from that population drawn by control number.

Cohen's d when an entire population, and not just a sample, is being analyzed.²¹

Notably, Commerce itself has not identified any texts that support its proposed use of Cohen's d . The only academic text cited by Commerce in support of its position is Professor Ellis's statement that "The best way to measure an effect is to conduct a census of an entire population...."²² But that statement does not indicate that Cohen's d should be used when measuring the effect on an entire population. After all, Cohen's d is not the only possible measure of effect size. Professor Cohen's text itself identifies eight different measures of effect size — labeled d , r , q , g , h , w , f , and f^2 — each of which is intended for use only with a specific test of statistical significance.²³ Nothing in the passage Commerce has quoted from Professor Ellis's text suggests that the passage was referring specifically to the d statistic, or that d is the only possible measure of effect size, or that d can properly be used

²¹ To be precise, Professor Cohen's text does envision the use of *estimates* of d for a population *ex ante* to design experiments. See, e.g., Cohen, STATISTICAL POWER ANALYSIS FOR THE BEHAVIORAL SCIENCES (2d ed. 1988) at 12, 27, 52, 66 (Appx3756, 3770, 3795, 3809). But he nowhere suggests that d can actually be measured for an entire population or that, when a measurement is made for an entire population, the normal requirements of Normality, equal variances, and sufficient data can be waived.

²² See Redetermination at 13 and 49 (Appx0040, 0076).

²³ See Cohen at vii-ix and 13 (2d ed.1988) (Appx3732-3734, Appx3757).

when conducting a census of an entire population that does not satisfy the assumptions described by Professor Cohen for the use of the d -statistic.

In these circumstances, Commerce’s argument is purely negative: It contends that none of the texts on the record prohibit its proposed use of Cohen’s d . And, it faults SeAH for failing to identify any texts that explicitly address Commerce’s argument that Cohen’s d can properly be used to analyze an entire population, even when the assumptions described by Professor Cohen are not satisfied.²⁴

As a legal matter, there is no basis for Commerce’s claim that SeAH failed to meet its burden — because Commerce, not SeAH, bears the burden of proof on this issue.²⁵ More importantly, the apparent silence of academic

²⁴ See Redetermination at 40 and 52 (Appx0067, 0079).

²⁵ The statute grants Commerce authority to depart from the normal A-to-A methodology only when certain factual prerequisites are satisfied. See 19 U.S.C. § 1677f-1(d)(1)(B). In such circumstances, the ultimate burden of proof for establishing the existence of those prerequisites falls on Commerce. See, e.g., *Creswell Trading Co. Inc. v. U.S.*, 15 F.3d 1054, 1060–61 (Fed. Cir. 1994) (“The ‘if’ clause ... sets forth on its face a statutory condition that Commerce must establish before it may exercise its right to levy a countervailing duty against an investigated party, as opposed to an exception into which that party must prove its actions fall. The ultimate burden of proof is thus upon Commerce to establish by a preponderance of the evidence” that the statutory condition has been satisfied.”).

texts with respect to Commerce’s argument does not, in fact, mean that Commerce’s proposed use conforms to widely-adopted statistical practices.

It is well-settled that Cohen’s d is an example of what is known in statistics as a “parametric test,” because its results depend only on the parameters of the mean and standard deviation of the data. Such tests are, by their nature, appropriate only when the distribution being analyzed is also fully described by those parameters. As a recent article explains,

The most commonly reported estimate for δ is Cohen’s d . Cohen’s d is one of a family of standardized mean difference (SMD) statistics ... whose purpose is to estimate δ by standardizing a treatment effect with either the pooled standard deviation (Cohen’s d).... *As parametric ES statistics, however, their accuracy in estimating δ depends substantially on distributional assumptions (i.e., normality, variance homogeneity) being met.*²⁶

²⁶ Ricca and Blaine, *Notes on a Nonparametric Estimate of Effect Size*, 90:1 JOURNAL OF EXPERIMENTAL EDUCATION 249 (2022) (emphasis added) (Appx6391).

This point is confirmed by other texts. For example, the Oxford Handbook of Quantitative Methods explains that:

Classic parametric statistics are the dominant method for analyzing data in psychology and related fields. Researchers routinely ... compute effect sizes such as Cohen’s d . There are important assumptions underlying classic parametric statistics—for example, that scores are normally distributed in the population.... Many psychologists do not understand that using classic parametric methods when the assumptions underlying them are sufficiently violated can lead to

(footnote continued on following page)

Furthermore, it is also well-established that the d -statistic — and the numerical thresholds for evaluating d that Commerce has used — were proposed by Professor Cohen solely for the purpose of measuring the “power” of an experiment in certain specified circumstances. The text in which Professor Cohen introduced his d -statistic actually includes numerous alternate measures of effect size, each of which is designed for use with a specific type of test of statistical significance.²⁷ According to that text, the d statistic was intended for use *solely* in conjunction with a “t-test.”²⁸ Professor Cohen identified seven different measures of effect size — labeled r , q , g , h , w , f , and f^2 — for use in other situations.²⁹

(footnote continued from previous page)

undesirable consequences.... These problems can lead to erroneous research findings.

1 *Oxford Handbook of Quantitative Methods* 388 (2013) (Appx6395).

²⁷ See Cohen at 13 (“Each of the Chapters 2-10 will present in some detail the ES index appropriate to the test to which the chapter is devoted.”) (Appx3757).

²⁸ A “t-test” calculates a “t” value by dividing the difference in means of two datasets by the “standard error” of the data. The standard error, in turn, is calculated by dividing the standard deviation of each dataset by the number of data points in the dataset. See Karris, Steven, *MATHEMATICS FOR BUSINESS, SCIENCE AND TECHNOLOGY* 10-72 to 10-74 (3d ed. 2007) (A copy of this text was provided in Attachment 11 of SeAH’s November 12 Submission.) (Appx5060-5062).

²⁹ See, e.g., Cohen at vii to ix (Appx3732-3734).

Commerce itself admits that a “t-test” can appropriately be used only when the data being analyzed satisfies the assumptions of a Normal distribution with equal variances.³⁰ Consequently, it is hardly surprising that academic texts do not directly address the use of Cohen’s *d* in circumstances in which statistical significance is not being measured by a t-test and in which the assumptions required for a t-test are not satisfied. Commerce’s proposed use of Cohen’s *d* to measure the “effect size” for data that, according to Commerce, encompasses an entire population has no relationship to the use of Cohen’s *d* in statistical practice.

2. Commerce Has Not Identified Any Logical or Factual Support for Its Claimed Use of Cohen’s d when Professor Cohen’s Assumptions Are Not Satisfied

It is a fundamental principle of statistical analysis that Normal and non-Normal distributions have very different mathematical properties.³¹ Among other things, Normal distributions are described completely by their mean and

³⁰ See Redetermination at 42-43 (Appx0069-0070).

³¹ See, e.g., Starnes, Yates, and Moore, *Statistics through Applications* 116 (2005). (A copy of this text was provided in Attachment 12 of SeAH’s November 12, 2021, Submission.) (Appx5371) (“Normal distributions play a large role in statistics, but they are rather special and not at all “normal” in the sense of being average or natural.”). See also SeAH’s Comments on Draft Redetermination at 13-17 (Appx5569-5573).

standard deviation.³² As a result, a parametric test (like Cohen's d) that considers only the mean and standard deviation of the data can appropriately be used for Normal distributions. Furthermore, if the data is known to be Normal, it is possible to calculate the number of data-points that will fall within a given number of standard deviations of the mean, without knowing anything more about the data.³³ Professor Cohen relied on that characteristic of Normal distributions when calculating the practical significance of his proposed thresholds for evaluating the d statistic.³⁴ But such calculations are not possible when the data is not Normal, and the distribution is not defined completely by its mean and standard deviation.

³² See, e.g., Starnes, Yates, and Moore, *Statistics through Applications* 135 (2005). (A copy of this text was provided in Attachment 12 of SeAH's November 12, 2021, Submission.) ("A specific Normal curve is completely described by its mean and standard deviation.") (Appx5390).

³³ In a Normal distribution, the well-known "empirical rule" states that 68.27 percent of data in a Normal distribution will fall within one standard deviation of the mean. But one can also calculate that 57.6 percent of the data will fall within 0.8 standard deviations of the mean, that 38.3 percent will fall within 0.5 standard deviations of the mean, and that 15.9 percent will fall within 0.2 standard deviations of the mean.

By contrast, if the data is not known to follow a Normal distribution, it is not possible to make such calculations. In a non-Normal distribution, it is possible that none of the data will fall within one standard deviation of the mean. It is also possible that virtually all of the data will fall within one standard deviation.

³⁴ See, e.g., Cohen at 25-27 (Appx3768-3770).

In light of the mathematical differences between Normal and non-Normal distributions, there is no reason to believe that a rule-of-thumb (such as Professor Cohen's proposed thresholds) that was developed to be used with Normal distributions can properly be applied to non-Normal data. Certainly, Commerce has not offered any mathematical calculations supporting the conclusion that Professor Cohen's proposed thresholds have the same meaning for non-Normal data that they would for Normal distributions.

In such circumstances, there is no basis for Commerce's claim that it is not required to consider the recognized limitations on the use of Cohen's d when it has the data for an entire population. Whether based on an entire population or derived from a sample, a Normal distribution has different mathematical characteristics than a non-Normal distribution. As a result, conclusions drawn about the meaning of a difference of 0.8 standard deviations for a Normal distribution that is completely defined by its standard deviation say nothing about the meaning of a difference of 0.8 standard deviations in a distribution that is not Normal.³⁵

³⁵ Chebyshev's inequality holds, in general that, in an arbitrary distribution, no more than $1/k^2$ of the data may be located more than k standard deviations from the mean. *See, e.g.,* Karris, MATHEMATICS FOR BUSINESS, SCIENCE AND TECHNOLOGY at 10-28 (formula 10.108). (A copy of this text was provided in Attachment 11 of SeAH's November 12 Submission.) (Appx5016).

(footnote continued on following page)

*B. Commerce's Use of the Cohen's d Test
Can Classify Imperceptible Differences
as Significant, Just as this Court Predicted*

In *Stupp III*, this Court expressed a separate concern that Commerce's simplistic application of its Cohen's d test could generate arbitrary results when applied to data with a small number of observations or small price differences. To support this point, the Court provided a hypothetical example showing that, when the variances in the data are small, even tiny differences that had no practical significance could result in a "large" value for d ³⁶

*1. Commerce's Claim that the "Meaningful Difference"
Test Takes Care of the Situations Described by the
Court's Hypothetical Example Is Incorrect When
There Is More than One Product Being Analyzed*

Commerce's Redetermination does not appear to dispute the conclusion that the example set forth in the *Stupp III* decision would generate a false-positive "passing" result under the Cohen's d test. However, the Redetermination asserted that any harm caused by that false positive would be ameliorated by the "meaningful difference" test, since the minor price differences generated by the *Stupp III* decision's hypothetical example would

(footnote continued from previous page)

For any value of k that is less than or equal to 1, Chebyshev's inequality does not generate meaningful results (since the value of $1/k^2$ will in such cases always be greater than or equal to 100 percent).

³⁶ See *Stupp III*, 5 F.4th at 1358-59.

result in an insignificant difference in the dumping margins under the different comparison methodologies.³⁷ In mathematical terms, if there are only tiny differences in the prices for individual U.S. sales of a specific product, the transaction-specific prices for that product will be similar to the average of the U.S. prices, and there should be little difference whether the transaction-specific prices or the average is used in the dumping calculations.

But Commerce's explanation holds true only if there is only one product under consideration. If there are multiple products under consideration, an incorrect finding that there are "large" price differences for any one product will affect the ultimate outcome of Commerce's "Ratio Test," which depends on the percentage of sales with "large" price differences. Consequently, even if an incorrect finding of a "large" price difference for one product does not affect the dumping margin calculated for that product, it may affect the results of the "Ratio Test" and, hence, the dumping margins calculated for other products.

It is a fairly simple matter to construct an example showing how, when there is more than one product under consideration, an incorrect finding of a "large" price difference for one product can affect the dumping margin

³⁷ Redetermination at 31 (Appx0058).

calculated for a different product for which there is no U.S. price difference at all. Consider, for example, a situation in which a seller sells two products to two U.S. customers. For the first product, it makes a total of eight sales, four sales to each customer, at an intended price of \$100. Suppose, further, that, due to random fluctuations, the net price for one of the sales turns out to be \$99.999, instead of \$100.³⁸ The following table summarizes the resulting prices and the outcome of the DPA for that product:

Product 1 Sales

| | Customer 1 | Customer 2 | Total |
|-------------------------|---------------|---------------|----------------|
| Sale 1 | 99.999 | 100.000 | |
| Sale 2 | 100.000 | 100.000 | |
| Sale 3 | 100.000 | 100.000 | |
| Sale 4 | 100.000 | 100.000 | |
| Total | 399.999 | 400.000 | |
| Average | 99.99975 | 100.000 | |
| Difference in Means | | | 0.00025 |
| Standard Deviation | 0.0047 | 0.0000 | 0.00031 |
| Cohen's <i>d</i> | | | 0.81650 |

Under the DPA's analysis, the Cohen's *d* for the comparison of the sales to the two customers would be found to be "large" (*i.e.*, greater than 0.8), and

³⁸ Such a difference might occur, for example, because of minor fluctuations in the exchange rate used to convert movement expenses incurred in a foreign currency. Or, they might arise from slight differences in the credit periods for individual sales with identical payment terms, where the payment term for one sale falls on a weekend and therefore is extended an extra day. Or, they might arise from any number of other random occurrences.

the analysis of the sales of this product to both customers would be found to “pass” the DPA’s Cohen’s d test. In other words, an insignificant difference in the price of one sale would result in a finding that the sales “passed” the Cohen’s d test for these customers, even though there is no “practical significance” to the difference in the sales prices.

Now, suppose that the seller also makes four sales of its second product to the same two customers, two sales to each, at prices of 90 for the first sale to each customer and 110 for the second sale to each customer. The following table summarizes the resulting prices and the outcome of the DPA:

Product 2 Sales

| | Customer 1 | Customer 2 | Total |
|-------------------------------|---------------|---------------|--------------|
| Sale 1 | 90.000 | 90.000 | |
| Sale 2 | 110.000 | 110.000 | |
| Total | 200.000 | 200.000 | |
| Average | 100.000 | 100.000 | |
| Difference in Means | | | 0.000 |
| Standard Deviation | 10.000 | 10.000 | 10.000 |
| Cohen’s d | | | 0.000 |

Under the DPA’s analysis, none of the sales of the second product to these customers would “pass” the Cohen’s d test, because the average price to the two customers is the same, and the resulting Cohen’s d is zero. However, when considering the seller’s overall U.S. sales of the two products, 66.7 percent (that is, the 8 sales of Product 1 out of the total of 12 sales of the two

products) would “pass” the Cohen’s *d* test. As a result, Commerce would apply the A-to-T methodology for all sales.

Now, suppose that the Normal Value for both products is 101.97. As illustrated in the following table, the overall average dumping margin for Products 1 and 2 *under the A-to-A methodology* would be *1.97 percent*.

Calculation of Dumping Margin for Products 1 and 2
Using Average-to-Average Comparison

| Product | Quantity (A) | Total U.S. Sales Value (B) | Average U.S. Price (C=B/A) | Normal Value (D) | Per-Unit Dumping Margin (E=D-C) | Extended Dumping Margin (F=AxE) | Percentage Dumping Margin (G=F/B) |
|-----------|-----------------|-------------------------------------|----------------------------------|------------------------|--|--|--|
| Product 1 | 8 | 799.999 | 99.9999 | 101.9700 | 1.9701 | 15.76 | 1.97% |
| Product 2 | 4 | 400.000 | 100.0000 | 101.9700 | 1.9700 | 7.88 | 1.97% |
| Total | 12 | 1,199.999 | | | | 23.64 | 1.97% |

By contrast, *under the A-to-T methodology*, the overall average dumping margin would be *3.31 percent*, as shown in the following table:

Calculation of Dumping Margin for Products 1 and 2
Using Average-to-Transaction Comparison

| Product | Sale | U.S. Price (A) | Normal Value (B) | Dumping Margin Amount (C=B-A) | Percentage Dumping Margin (D=C/A) |
|-------------|-------------------|----------------------|------------------------|--|--|
| Product 1 | Customer 1 Sale 1 | 99.999 | 101.97 | 1.9701 | |
| | Customer 1 Sale 2 | 100.000 | 101.97 | 1.9700 | |
| | Customer 1 Sale 3 | 100.000 | 101.97 | 1.9700 | |
| | Customer 1 Sale 4 | 100.000 | 101.97 | 1.9700 | |
| | Customer 2 Sale 1 | 100.000 | 101.97 | 1.9700 | |
| | Customer 2 Sale 2 | 100.000 | 101.97 | 1.9700 | |
| | Customer 2 Sale 3 | 100.000 | 101.97 | 1.9700 | |
| | Customer 2 Sale 4 | 100.000 | 101.97 | 1.9700 | |
| Subtotal | | 799.999 | 815.76 | 15.7601 | 1.97% |
| Product 2 | Customer 1 Sale 1 | 90.00 | 101.97 | 11.9700 | |
| | Customer 1 Sale 2 | 110.00 | 101.97 | 0.0000 | |
| | Customer 2 Sale 1 | 90.00 | 101.97 | 11.9700 | |
| | Customer 2 Sale 2 | 110.00 | 101.97 | 0.0000 | |
| Subtotal | | 400.00 | 407.88 | 23.9400 | 5.99% |
| Grand Total | | 1,199.999 | 1,223.64 | 39.7001 | 3.31% |

In short, even if a false finding of a “large” price difference for the first product does not result in a finding of dumping for that individual product, it may change the methodology used to calculate the dumping margin for another product, and thus create an overall finding that dumping exists. The “meaningful difference” test will not prevent the unfair creation of dumping margins in such cases. This Court was, therefore, correct when it suggested that Commerce’s improper use of Cohen’s *d* could improperly create

affirmative determinations of dumping, when the A-to-A methodology would indicate that only *de minimis* dumping existed.

Surprisingly, the CIT found that the flaw this Court identified in Commerce’s methodology did not mean that Commerce’s methodology was unreasonable. According to the CIT,

It is reasonably discernable that Commerce does not rely on the meaningful difference test to prevent all “inappropriate” passes from affecting a respondent’s dumping margins. Commerce has explained the meaningful difference test compensates for a specific concern with low-variance sales which the Court of Appeals identified.... SeAH’s argument is misplaced, because the question before the court is not whether it is possible to construct an unusual scenario where Cohen’s d test can result in an alternative comparison method. Rather, the question is whether Commerce’s use of Cohen’s test, when applied as a component of its differential pricing analysis, is reasonable.³⁹

According to the CIT, the calculations provided by SeAH were merely “hypothetical,” and did not show an actual error in the analysis of SeAH’s prices.⁴⁰

But the “unusual scenario” relied upon by SeAH was a slight modification of a hypothetical scenario that *this Court* asked Commerce to address — where prices with a tiny variance differed on average by an

³⁹ See *Stupp IV*, 619 F.Supp.3d at 1328.

⁴⁰ *Id.*, at n.13.

imperceptible amount. As SeAH demonstrated, Commerce’s response to the scenario presented by this Court failed to address was based on a simplification that failed to address the Court’s underlying concern. The CIT’s dismissal of SeAH’s demonstration as a mere “hypothetical” improperly dismisses *this Court’s* concern and ignores the fundamental issue.

2. *Commerce’s “Ratio Test” Cannot Give Meaningful Results when the Cohen’s d Test Fails to Properly Identify Significant Price Differences*

The failure of the “meaningful difference” test described above highlights a fundamental error in one of the arguments relied upon by Commerce’s Redetermination: Contrary to Commerce’s claims, the individual elements of Commerce’s Differential Pricing Analysis are not independent.⁴¹ Instead, they operate as a whole to implement the statutory directive. A failure of any one piece topples the entire edifice.⁴²

⁴¹ *Cf.* Redetermination at 41 and 42 (Appx0068-0069).

⁴² In this regard, the CIT’s decision appears to be based on an irreconcilable internal contradiction. In upholding Commerce’s DPA against the claim that it fails to distinguish true patterns from chance fluctuations, the CIT asserts that the “pattern” requirement is addressed only and entirely by Commerce’s “Ratio Test,” and not by the Cohen’s *d* test. *See Stupp IV*, 619 F.Supp.3d at 1327. But, when upholding the DPA against the claim that the Cohen’s *d* test may mistakenly find minor, random price fluctuations to be “significant” when the conditions required by Professor Cohen are not satisfied, the CIT asserts that the DPA must be evaluated as an integrated whole. *Id.* at 1325
(footnote continued on following page)

Commerce has made clear that, in order to make a finding that a “pattern” exists that would justify a departure from the normal A-to-A comparison under the statute, it must distinguish between differences that have meaning and those that arise by chance. As Commerce has explained,

a ‘pattern of prices that differ significantly among purchasers, regions or time periods’ means that the Department is examining the extent to which the prices, when ordered by purchaser, region or time period, exhibit differences which have meaning, which have or may have influence or effect, which are noticeably or measurably large, and *which may be beyond something that occurs by chance....*”⁴³

Commerce has asserted, and the CIT appears to have agreed, that this “pattern” requirement is addressed only and exclusively by the DPA’s “Ratio Test,” and that this Court’s decision in *Stupp III* upheld the reasonableness of

(footnote continued from previous page)

(“Thus, even if the Cohen’s d values of small test groups were less accurate than for large test groups, this difference does not by itself render Commerce’s use of Cohen’s test unreasonable, because the ratio test and meaningful difference test compensate for inaccuracies.”). *See also id.* at 1326 (“Commerce addresses these arguments by explaining that even if Cohen’s test can produce positive results under unusual circumstances, this possibility does not mean its use of Cohen’s d is unreasonable when combined with the ratio test and meaningful difference test.”).

⁴³ *See* Issues and Decision Memorandum for the Final Affirmative Determination in the Less-Than-Fair-Value Investigation of Welded Line Pipe from the Republic of Korea (Oct. 5, 2015) (“Final I&D Memo”) at 20 (emphasis added) (Appx0238).

the “Ratio Test” for that purpose. In our view, that claim represents a gross mis-reading of the *Stupp III* decision. In fact, it was never argued in that proceeding, and the *Stupp III* decision never claimed that the “Ratio Test” by itself reasonably distinguished between true patterns and chance fluctuations.⁴⁴

⁴⁴ This Court’s decision in *Stupp III* held that the 33- and 66-percent cut-offs selected by Commerce represented a reasonable *policy* choice:

Commerce’s selection of the 33% and 66% cutoffs is a reasonable choice. An alternative approach might be, for example, to use a single cutoff at 50%. That approach would undoubtedly favor some respondents—the more frequent application of the average-to-average method would result in more de minimis dumping margins—but it would disfavor other respondents. For example, respondents having slightly more than 50% of their sales passing the Cohen’s d test would have the average-to-transaction method applied to all of their sales. Commerce’s approach is less rigid, providing a middle ground between 33% and 66%, in which the average-to-transaction method is only partially applied. That approach provides a better fit, minimizing both the assessment of antidumping duties that are too high and the assessment of duties that are too low. We conclude that Commerce’s cutoffs are reasonable in light of the alternatives.

See Stupp III, 5 F.4th at 1355. Nothing in the Court’s reasoning states that the 33- and 66-percent thresholds represent a reasonable tool for distinguishing true patterns from random fluctuations. Such a conclusion would require a mathematical analysis of the likelihood that a ratio of 33 or 66 percent is unlikely to have occurred by chance, not general statements about how different cut-offs might affect respondents.

But, no matter how *Stupp III* is read, it is clear that neither the “Ratio Test” in particular, nor the DPA as a whole, can properly distinguish between true patterns and chance fluctuations when the “passing” scores counted by the “Ratio Test” are themselves meaningless. In other words, if the results of the Cohen’s *d* test are meaningless, then counting those results for purposes of the “Ratio Test” will only give meaningless figures that cannot reasonably provide the basis for any decision.

Importantly, when Cohen’s *d* is properly used, it *does* provide a basis for distinguishing between true patterns and chance fluctuations. Indeed, Professor Cohen proposed his “power” analysis using Cohen’s *d* precisely for that purpose.⁴⁵ As Dr. Ellis has explained, “Statistical power describes the probability that a test will correctly identify a genuine effect.”⁴⁶ And, as Professor Cohen demonstrated, the statistical power of any test can be

⁴⁵ The first paragraphs of Professor Cohen’s textbook explained that,

The power of a statistical test is the probability that it will yield statistically significant results... The purpose of this book is to provide a self-contained comprehensive treatment of statistical power analysis from an “applied” viewpoint.

Cohen at 1 (Appx3745).

⁴⁶ See Ellis, Paul, THE ESSENTIAL GUIDE TO EFFECT SIZES: STATISTICAL POWER, META-ANALYSIS, AND THE INTERPRETATION OF RESEARCH RESULTS (2010) at 52 (Appx4436).

calculated (as long as the required criteria are satisfied) solely from the value of d and the number of data-points.⁴⁷ Crucially, these calculations were not based on “real-world” observations about heights or IQs, but were, instead, derived solely from mathematical analysis. That analysis, in turn, explicitly requires the assumption that the data was drawn from Normal distributions and had roughly equal variances.⁴⁸

Consequently, when the criteria for the proper application of Cohen’s d are satisfied, an analysis of the number of sales with a “large” d value may, indeed, allow Commerce to distinguish between true patterns and chance fluctuations.⁴⁹ But, that conclusion is possible only when the mathematical

⁴⁷ In tables set forth in his text, Professor Cohen calculated the likelihood that any result was statistically significant, based on the number of observations, the effect size (either assumed or measured), and the desired level of statistical significance. According to his calculations, if the value of d is 0.8, two data sets each containing 26 observations would be needed to establish, with a likelihood of 80 percent, that the observed phenomenon was statistically significant at the 95-percent level normally required by statisticians. *See* Cohen at 36 (Appx3779).

⁴⁸ *See* Cohen at 27 (“The power tables are designed to yield power values for the t test for the difference between the means of two independent samples of equal size drawn from normal populations having equal variances.”) (Appx3770).

⁴⁹ As we explained to this Court in the proceedings leading to the *Stupp III* decision, Commerce has not attempted to show that the numerical cut-offs adopted in the “Ratio Test” would achieve that purpose. Thus, while
(footnote continued on following page)

assumptions underlying Professor Cohen’s calculations are satisfied. A failure to apply Cohen’s d in a mathematically-correct manner necessarily precludes reliance on the “Ratio Test” to distinguish between true patterns and chance fluctuations.

C. Commerce’ Attempt to Show that Its Use of the Differential Pricing Analysis Affected the Outcomes in Only a Small Number of Investigations Misrepresents the Actual Impact of Commerce’s Methodology and Is Fundamentally Irrelevant

Commerce’s Redetermination includes an analysis of the results of all investigations during 2015 and 2021, which purports to show that the application of the DPA resulted in the use of an alternate comparison methodology in only 21 or 22 percent of its investigations.⁵⁰ The relevance of this analysis is questionable: As Commerce conceded, “the results found for other companies in other LTFV investigations have no bearing on the results found for SeAH in the Final Determination.”⁵¹ Nevertheless, Commerce asserted that, “the broad picture of Commerce’s determinations supports the earlier conclusion that the Differential Pricing Analysis is reasonable in that it

(footnote continued from previous page)

some analysis based on the number of sales passing the Cohen’s d might be reasonable, there is no reason to believe that the “Ratio Test” is.

⁵⁰ See Redetermination at 32 (Appx0059).

⁵¹ *Id.* at 60 (Appx0087).

compares the frequency of where an alternative comparison methodology is applied to the frequency of where the standard comparison methodology is applied.”⁵² The CIT’s finding that Commerce’s use of the Cohen’s *d* test was reasonable appears to have been based, in large part, on the conclusion that a finding of “patterns” in 21 or 22 percent of cases appeared reasonable on its face.⁵³

But Commerce’s argument, and the CIT’s analysis, are fundamentally illogical. The statute does not direct Commerce to develop a methodology that would lead to a change in comparison methodologies in a reasonable number of cases. Instead, it authorizes Commerce to depart from the normal A-to-A methodology only when specific factual criteria were satisfied. In such circumstances, a finding of reasonableness requires a comparison of the DPA’s results with the number of cases in which the factual criteria described

⁵² *Id.* (Appx0087).

⁵³ *See Stupp IV*, 619 F.Supp.3d at 1327 (“Discernible from Commerce’s explanation is that the 0.8 cutoff produces reasonable passing rates once the ratio and meaningful difference tests are applied.... Commerce’s decision to adopt Cohen’s 0.8 (“large”) threshold as a measure of significance because it is widely accepted in the statistical literature does not undermine the reasonableness of that choice, if it is based on Commerce’s expertise and Commerce demonstrates the reasonableness of that choice with reference to the impact it has on the differential pricing analysis. Thus, Commerce’s reference to Cohen’s work does not circumscribe its discretion to choose the same values in a new context, because that choice is itself reasonable.”).

by the statute are actually satisfied. If the factual criteria are satisfied in only 5 percent of cases, then the fact that Commerce's methodology results in alternate methodologies in 21 or 22 percent should be disturbing. By the same token, if the factual criteria are satisfied in 50 percent of cases, then the fact that Commerce's methodology results in alternate methodologies in 21 or 22 percent would be disturbing from a different direction. It is only reasonable for Commerce's methodology to result in alternate methodologies in 21 or 22 percent of cases if the actual number of cases satisfying the factual criteria described in the statute is close to 21 or 22 percent. Without knowing what the "correct" result should be, there is no way to tell whether Commerce's actual results are reasonable or not.

Furthermore, Commerce's statistics about the impact of the DPA on its results, which were relied upon by the CIT, are themselves misleading. In particular, Commerce's statistics are skewed by the inclusion of investigations where the dumping margins would have been above *de minimis* regardless of whether the A-to-A and A-to-T methodology was used to calculate the margin. In this regard, the standards applied by Commerce differ when the A-to-A margins are above *de minimis* are different from the standards used when the dumping margins would be *de minimis* under the A-to-A methodology: When the dumping margins are above *de minimis*,

Commerce finds a “meaningful difference” in the results only when the A-to-T margins would be 25 percent higher than the A-to-A margins. By contrast, when the A-to-A margins would be *de minimis*, any increase in the margins to above *de minimis* is considered a “meaningful difference.”

In SeAH’s case, the dumping margins under the A-to-A methodology would have been a *de minimis* 1.97 percent. Under Commerce’s methodology, an increase in SeAH’s margin of 0.03 percentage points, from 1.97 percent to 2.00 percent, would have been a “meaningful difference.” That is a very different situation from cases in which the dumping margin is already above *de minimis*, and a further increase of 25 percent in the dumping margin is needed to establish a “meaningful difference.”

In our comments on Commerce’s Redetermination, we presented an analysis of all of Commerce’s investigations between 2015 and 2021 in which the dumping margins would have been *de minimis* under the A-to-A methodology.⁵⁴ The following table summarizes the results.

⁵⁴ See SeAH’s Comments on Final Redetermination (June 14, 2022) at 35 and Attachment 3 (Appx5638, 5661-5669).

Results of Investigations that Would Have Had
de Minimis Margins under A-to-A Comparison
2015 to 2021

| Year | Negative Final Determination (A) | Affirmative Final Using A-to-T or Mixed (B) | Total Investigations (C=A+B) | Percentage Affirmative Due to DPA (D=B/C) |
|-------|----------------------------------|---|------------------------------|---|
| 2015 | 2 | 3 | 5 | 60.00% |
| 2016 | 4 | 5 | 9 | 55.56% |
| 2017 | 1 | 9 | 10 | 90.00% |
| 2018 | 5 | 1 | 6 | 16.67% |
| 2019 | 2 | 2 | 4 | 50.00% |
| 2020 | 6 | 5 | 11 | 45.45% |
| 2021 | 8 | 6 | 14 | 42.86% |
| Total | 28 | 31 | 59 | 52.54% |

As this analysis demonstrates, Commerce’s use of the DPA transformed *de minimis* dumping margins into affirmative determinations of dumping in more than 50 percent of the investigations in which there would have been a negative determination under the A-to-A methodology. That is very different from Commerce’s claim that its “differential pricing analysis” only changed the results in 21 or 22 percent of cases.

In these circumstances, Judge Kelly’s reliance on Commerce’s 21 or 22 percent figure to find Commerce’s practice “reasonable” has absolutely no relationship to the actual situation in this case. This is not a proceeding where the A-to-A margins were already above *de minimis*, and an increase in the margin of 25 percent was needed to create a “meaningful difference” under Commerce’s practice. Instead, SeAH’s dumping margin under an A-to-A comparison would have been *de minimis*, and all that was required to

constitute a “meaningful difference” under Commerce’s practice was an increase in the margin of 0.03 percentage points (which would have been 1.5 percent of the A-to-A margin). In such circumstances, the likelihood that a dumping margin would be created by Commerce’s DPA was more than 50 percent. While such figures cannot be evaluated in a vacuum, it seems doubtful that any objective observer would consider such an outcome on its face to be “reasonable.”

D. Commerce’s Assertion that Professor Cohen’s Proposed Thresholds Can Be Used as Universal Yardsticks Because They Are Based on Real-World Observations Is Illogical and Contrary to the Evidence

As described above, Commerce’s Redetermination did not attempt to justify its use of Professor Cohen’s proposed thresholds as a matter of statistical practice or as a matter of mathematical logic. Instead, Commerce fell back on the claim that Professor Cohen’s rule-of-thumb that a d of 0.8 indicates a “large” effect can be used as a universal yardstick, because it reflects “real-world observed differences.”⁵⁵ Indeed, Commerce’s Redetermination concludes with the observation that “Dr. Cohen’s thresholds

⁵⁵ See Redetermination at 20, 24, 28 (Appx0047, 0051, 0055).

are not based on the alleged statistical criteria but, rather, on real-world observations.”⁵⁶

It is, undoubtedly, true that Commerce is not required to follow statistical principles in its dumping calculations. But that does not mean that it is reasonable for Commerce to rely on any rule-of-thumb that is purportedly derived from real-world observations of some unrelated phenomenon. Instead, before Commerce can rely on such a rule-of-thumb, it must demonstrate that the phenomenon from which the rule-of-thumb was derived has some rational relationship to the situation for which Commerce intends to apply that rule. In this case, Commerce has absolutely failed to meet that burden.

*1. Professor Cohen Expressly Rejected
the Notion that His Rules-of-Thumb
Established Universal Yardsticks*

As mentioned, Professor Cohen supported his proposed thresholds by reference to the differences in average heights of teenaged girls of different ages and on the differences in average IQs of different types of students.⁵⁷ He

⁵⁶ *Id.* at 65 (Appx0092).

⁵⁷ *See* Cohen at 27 (Appx3770).

admitted that the thresholds he selected were arbitrary,⁵⁸ and that his selection process was inherently imprecise.⁵⁹ As he subsequently explained,

My intent was that medium ES represent an effect likely to be visible to the naked eye of a careful observer. I set small ES to be noticeably smaller than medium but not so small as to be trivial, and I set large ES to be the same distance above medium as small was below it.⁶⁰

In his text, Professor Cohen described his “large” threshold as identifying effects that were “grossly perceptible.”⁶¹

At the same time, Professor Cohen made clear that his proposed rules-of-thumb were not intended as a universal yardstick for evaluating effect sizes. His text emphasized that an attempt to derive a single measure of effect size would be “self-defeating.”⁶² And, he specifically cautioned against the use of

⁵⁸ See Cohen at 12 (Appx3756).

⁵⁹ See J. Cohen, *A Power Primer*, 112:1 PSYCHOLOGICAL BULLETIN 155, 156 (1992) (Appx6388-6389). See also Ellis at 41 (2010) (Appx4425).

⁶⁰ J. Cohen, *A Power Primer* at 155, 156 (Appx6388-6389).

⁶¹ See Cohen at 27 (Appx3770).

⁶² As Professor Cohen explained,

From one point of view, a universal ES index, applicable to all the various research issues and statistical models used in their appraisal, would be the ideal. Apart from some formidable mathematical-statistical problems in the way, even if such an ideal could be achieved, the result would
(footnote continued on following page)

his rules-of-thumb as a uniform tool for assessing whether a given value of d was “small,” “medium,” or “large” measure even for research within different fields of behavioral science. To the contrary, he warned that such an assessment had to be “relative ... the area of behavioral science or even more particularly to the specific content and research method being employed in any given investigation.”⁶³ According to Professor Cohen, his proposed thresholds were intended “for use only when no better basis for estimating the {effect-size} index is available.”⁶⁴

In short, despite the fact that his rules-of-thumb were purportedly based on real-world phenomena, Professor Cohen himself denied that they were universally applicable even within the narrow realm of behavioral-science experiments. Commerce’s suggestion that those rules-of-thumb can nevertheless be used to assess calculated effect sizes for price differences is, therefore, inconsistent with Professor Cohen’s own understanding of the acceptable use of his proposed thresholds.

(footnote continued from previous page)

express ES in terms so unfamiliar to the researcher in behavioral science as to be self-defeating.

See Cohen at 11 (Appx3755).

⁶³ *Id.*, at 25 (Appx3768).

⁶⁴ *Id.* (Appx3768).

2. *There Is No Reason to Expect SeAH's U.S Prices to Follow a Similar Distribution to Heights or IQs*

As mentioned, Professor Cohen's text makes clear that the d statistic was intended to be used only when the datasets being compared met the criteria of Normal distributions, approximately equal variances, and sufficient data-points. Not surprisingly, then, the real-world examples that he used to support proposed thresholds for assessing whether a given d was small, medium, or large satisfied those criteria. Among other things, both the heights of teenage girls and IQ scores are Normally distributed.⁶⁵

All Normal distributions share fundamental mathematical characteristics. Indeed, "all Normal distributions *are the same* if we measure in units of size σ about the mean μ , as center."⁶⁶ Consequently, it may be mathematically reasonable to extend the thresholds that Professor Cohen found appropriate for Normally-distributed heights and IQs to other Normally-distributed data.

But that logic only applies to Normally-distributed data. It does not apply to non-Normal data. In this case, SeAH's pricing data did not follow a

⁶⁵ See Starnes, Yates, and Moore, at 124, 135 (2005). (A copy of this text was provided in Attachment 12 of SeAH's November 12, 2021, Submission.) (Appx5579, 5590).

⁶⁶ See *id.* at 123 (Appx5578) (emphasis added).

Normal distribution,⁶⁷ and it did not have equal variances or a sufficient number of data-points in the groups being compared.⁶⁸ Consequently, there is no mathematical reason to expect SeAH's prices to have the same characteristics as the data that Professor Cohen examined when selecting his proposed thresholds.

In the absence of a mathematical justification, Commerce must point to some reason for expecting SeAH's prices to have the same inherent

⁶⁷ See SeAH's September 9, 2015, Case Brief at 31, n.49. (Appx1311).

⁶⁸ For example, Commerce's final margin calculation printout shows that its Cohen's *d* test included comparisons of, *inter alia*:

- (1) a test group consisting of [^{Numbers}] transactions and a standard deviation of [^{Numbers}] to a base group consisting of [^{Numbers}] transactions with a standard deviation of [^{Numbers}];
- (2) a test group consisting of [^{Numbers}] transactions and a standard deviation of [^{Numbers}] to a base group consisting of [^{Numbers}] transactions with a standard deviation of [^{Numbers}];
- (3) a test group consisting of [^{Numbers}] transactions and a standard deviation of [^{Numbers}] to a base group consisting of [^{Numbers}] transactions with a standard deviation of [^{Numbers}]; and
- (4) a test group consisting of [^{Numbers}] transactions and a standard deviation of [^{Numbers}] to a base group consisting of [^{Numbers}] transactions with a standard deviation of [^{Numbers}].

For all of those comparisons, the Department found the sales to "pass" the Cohen's *d* test. See Memorandum concerning the "Final Determination Margin Calculation for SeAH" (Oct. 5, 2015), Attachment 2, at pages 113-15 (Appx1793-1795).

characteristics as the height and IQ data considered by Professor Cohen.

There is none: SeAH's prices are not based on the heights or IQs of its customers.⁶⁹ Consequently, the notion that SeAH's prices should be judged by rules-of-thumb designed for use with heights and IQs and other Normally-distributed data is absurd.

3. *The Evidence Confirms that There Are No Discernible Differences in the Prices for Products that, According to Commerce, Show a "Large" Effect Size under Its Cohen's d Test*

As mentioned, Professor Cohen's "medium" threshold was intended to identify differences that were "likely to be visible to the naked eye of a careful observer," and his "large" threshold was intended to identify effects that were "grossly perceptible."⁷⁰ A difference in prices that constitutes a "large" effect should, therefore, be "visible to the naked eye" and "grossly perceptible" when the data is examined. But examination of SeAH's actual

⁶⁹ See, e.g., SeAH's January 14, 2015, Section A Response at 26 ("The terms of sale and prices for sales of Line Pipe to customers in the home market and in the United States vary as a result of sale-by-sale negotiations. SeAH Steel and PPA do not have price lists setting different selling prices or terms of sale depending on the particular customer category or level of trade.") (Appx6336).

⁷⁰ See above at 12.

prices in this case demonstrates that Commerce has found a “large” effect even when the price differences are not visually discernible.

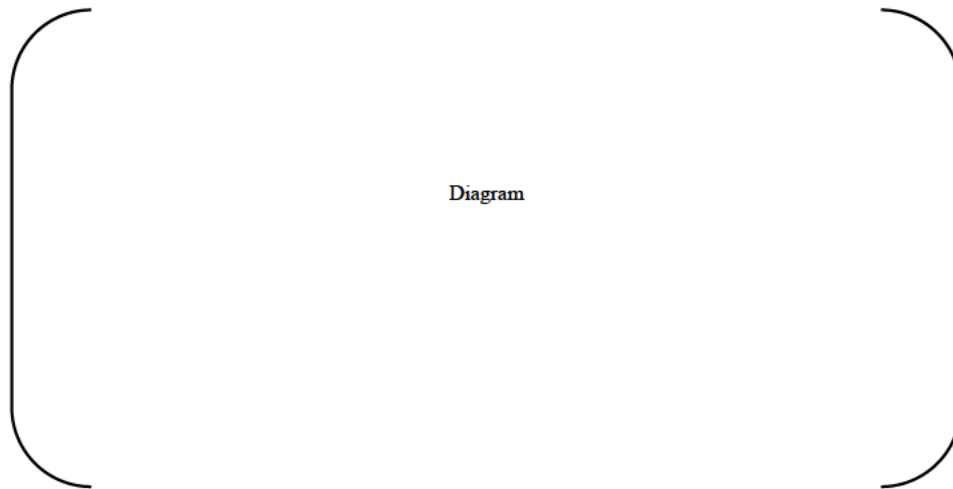
Consider, for example, the very first customer (number 102020) analyzed by Commerce’s margin calculation program, and the first product (control number 1-03-03-06-1) that was found to pass the Cohen’s d test for that customer.⁷¹ According to Commerce’s margin calculation program, SeAH had [Numbers] U.S. sales of control number 1-03-03-06-1 to customer 102020, with an average price of [Numbers], and a standard deviation of [Numbers]. For the “base group” consisting of U.S. sales of that product to all other customers, there were [Numbers] transactions, with an average price of [Numbers], and a standard deviation of [Numbers]. Based on this information, the margin calculation program calculated a difference in average prices of [Numbers], a pooled standard deviation of [Numbers], and a Cohen’s d of 1.00739. Because that d was above 0.8, the margin calculation program treated the sales of 1-03-03-06-1 to customer 102020 as “passing” the Cohen’s d test.

As discussed in detail above, Cohen’s d is based on the assumption that the sales of the two groups follow a Normal distribution, have roughly equal

⁷¹ The output of the Cohen’s d test for customer 102020 is found at pages 112 to 115 of the printout of the “U.S. Sales Margin Program” released by the Department in Attachment 2 of its October 5, 2015, Memorandum concerning the “Final Determination Margin Calculation for SeAH.” (Appx1792-1795).

variances (approximately equal to the pooled standard deviation), and have a roughly equal number of data-points. Graphically, two Normal distributions with the means identified by Commerce ([Numbers] for customer 102020 and [Numbers] for the base group), with variances equal to the pooled standard deviation of [Numbers] calculated by Commerce, and with a large enough number of data-points would look like this:

Normal Distributions with the
Same Means and Standard Deviations
as the “Target” and “Base” Groups for U.S. Sales
of Control Number 1-03-03-06-1 to Customer 102020



Visual examination of the diagram appears to show a “pattern” of price differences between the two groups, with the prices for customer 102020 appearing to be markedly *lower* than those of the base group.

However, this apparent pattern is an artifact of the assumption that the prices are Normally distributed and have equal variances. The actual dispersion of prices is quite different. The following diagram, which shows the actual individual prices for sales to the target and base groups (by date of sale), demonstrates that there is no clear pattern at all.

Actual Prices for U.S. Sales
in the "Target" and "Base" Groups for Sales
of Control Number 1-03-03-06-1 to Customer 102020



Diagram

Visual inspection of the actual individual sales prices in this diagram does not reveal any actual pattern by customer. The sales of control number

1-03-03-06-1 to customer 102020 are clustered precisely in the middle of the prices for sales of that product to other customers.⁷²

Strikingly, visual examination of the actual prices also demonstrates that the relationship suggested by Commerce's comparison of hypothetical Normal distributions is exactly backwards: As discussed above, comparison of the hypothetical Normal distributions suggested that the prices for sales of control number 1-03-03-06-1 to customer 102020 were *lower* than the prices for sales to other customers. But, the actual data shows that, in reality, the prices for sales to customer 102020 were *higher* than the prices for all sales to other customers made on the same date.⁷³

⁷² Numerical analysis of the actual prices reveals that there were 13 sales to the base group at prices below the lower of the two sales to customer 102020; 17 sales to the base group at prices above the higher of the two sales to customer 102020; and 6 sales to the base group at prices in-between the prices to customer 102020.

⁷³ Commerce's Redetermination asserts that a focus on sales to different customers on the same date improperly combines patterns by purchaser and time period. *See* Redetermination at 60 (Appx0087). But that argument is illogical: If SeAH makes its sales to all customers on the same date at the same price, but prices vary by date, then one might say that there is a pattern of price differences by time period. But there is no pattern of price differences by customer, since all customers receive the same price on the same date. Any apparent differences in average prices caused by the fact that different customers purchase on different dates is simply a reflection of the pattern of differences by time period, and not a real pattern of price differences by customer.

One can undertake similar analyses for other control numbers. In each case, it is clear that SeAH's sales data does not follow a Normal distribution, and that the stylized Normal distribution assumed by the DPA's Cohen's d test does not accurately reflect the actual distribution of prices or the relationships between the prices in the "test" and "base" groups being compared. In such circumstances, reliance on a d statistic calculated under false assumptions to find a pattern that is not actually discernible in the data cannot be reconciled with the evidence on the record.

When faced with this evidence, Commerce's Redetermination simply rejected the notion that a "large" effect should be apparent from visual examination of the data:

SeAH's support for the alleged "false-positive" result is unavailing. First, a "visual examination of the diagram," perhaps based on the construct that one will know it when one sees it, is inadequate when analyzing detailed data involving complex calculations.⁷⁴

But that position is inconsistent with Professor Cohen's explicit statements that his "medium" threshold should identify an effect "visible to the naked eye" and that his "large" threshold should identify an effect that is "grossly

⁷⁴ Redetermination at 59 (footnote omitted) (Appx0086).

perceptible.”⁷⁵ The evidence demonstrates that the actual differences in prices for SeAH’s sales are not “grossly perceptible,” and thus are not “large” under Professor Cohen’s standard, no matter what *d* value was calculated by Commerce.

CONCLUSION

In light of the foregoing, we respectfully request that the Court reverse the decision by the CIT and remand this case with instructions requiring Commerce to recalculate the dumping margins for SeAH using an average-to-average comparison.

Respectfully submitted,

/s/Jeffrey M. Winton

Jeffrey M. Winton
Amrietha Nellan
Jooyoun Jeong

WINTON & CHAPMAN PLLC
1100 13th Street, N.W., Suite 825
Washington, D.C. 20005
(202) 774-5500

Attorney for SeAH Steel Corporation

July 25, 2023

⁷⁵ See above at 44.

ADDENDUM

Stupp Corp. et al. v. United States,
Slip Op. 23-23 (CIT, Feb. 24, 2023)

Slip Op. 23-23

UNITED STATES COURT OF INTERNATIONAL TRADE

STUPP CORPORATION ET AL.,

**Plaintiffs and Consolidated
Plaintiffs,**

and

MAVERICK TUBE CORPORATION,

**Plaintiff-Intervenor and
Consolidated Plaintiff-Intervenor,**

v.

UNITED STATES,

Defendant,

and

**SEAH STEEL CORPORATION AND
HYUNDAI STEEL COMPANY,**

**Defendant-Intervenors and
Consolidated Defendant-
Intervenors.**

Before: Claire R. Kelly, Judge

Consol. Court No. 15-00334

OPINION AND ORDER

[Sustaining the U.S. Department of Commerce's third remand redetermination in the less-than-fair-value investigation of welded line pipe from the Republic of Korea.]

Dated: February 24, 2023

Jeffrey M. Winton and Jooyoun Jeong, Winton and Chapman PLLC, of Washington, D.C., argued for plaintiff SeAH Steel Corporation.

Consol. Court No. 15-00334

Page 2

Robert R. Kieपुरa, Trial Attorney, Commercial Litigation Branch, Civil Division, U.S. Department of Justice, of Washington, D.C., argued for defendant. With him on the brief were Claudia Burke, Assistant Director, Patricia M. McCarthy, Director, and Brian M. Boynton, Principle Deputy Assistant Attorney General. Of Counsel was Mykhaylo Gryzlov, Senior Counsel, Office of the Chief Counsel for Trade Enforcement and Compliance, U.S. Department of Commerce, of Washington, D.C.

Jeffrey D. Gerrish, Schagrin Associates, of Washington, D.C., argued for plaintiff Welspun Tubular LLC USA. With him on the brief were Roger B. Schagrin and Saad Y. Chalchal.

Kelly, Judge: Before the court is the U.S. Department of Commerce's ("Commerce") third remand redetermination in its 2015 less-than-fair-value investigation of welded line pipe imported from the Republic of Korea ("Korea"). See Final Results of Redetermination Purs. Ct. Remand, April 4, 2022, ECF No. 208 ("Remand Results"); see also Welded Line Pipe From [Korea], 80 Fed. Reg. 61,366 (Dep't Commerce Oct. 13, 2015) (final determination of sales at less than fair value), as amended by Welded Line Pipe From [Korea], 80 Fed. Reg. 69,637 (Dep't Commerce Nov. 10, 2015) ("Amended Final Determination") and accompanying Issues & Decisions Memo, A-580-876, (Oct. 5, 2015), ECF No. 30-3 ("Final Decision Memo"). In Stupp Corporation v. United States, the Court of Appeals for the Federal Circuit vacated this court's opinion, remanding to Commerce to further explain why it is reasonable to apply the Cohen's *d* test as part of its differential pricing analysis if certain statistical assumptions have not been met. Stupp Corporation v. United States, 5 F.4th 1341 (Fed. Cir. 2021) ("Stupp III"). For the following reasons, the court sustains Commerce's third remand redetermination.

BACKGROUND

The court presumes familiarity with the facts of this case as set out in this court's previous opinions, as well as the Court of Appeals' decision in Stupp III, and now recounts only the facts relevant to the court's review of the Remand Results. On November 14, 2014, Commerce initiated an antidumping duty investigation of welded line pipe from Korea. Welded Line Pipe From [Korea], 79 Fed. Reg. 68,213, 68,213 (Dep't Commerce Nov. 14, 2014) (initiation of less-than-fair-value investigation). Commerce published its final determination on October 5, 2015 and, finding that 39.72% of SeAH Steel Corporation's ("SeAH") U.S. sales passed the Cohen's *d* test, applied the average-to-transaction method to those sales. Final Decision Memo. at 4. Commerce accordingly calculated a 2.53% dumping margin for SeAH. Amended Final Determination at 69,638. SeAH appealed, arguing that Commerce's differential pricing analysis and application of the Cohen's *d* test were contrary to law and unsupported by substantial evidence. See Stupp Corp. v. United States, 359 F. Supp. 3d 1293, 1302 (Ct. Int'l Tr. 2019) ("Stupp I"), reconsideration denied, 365 F. Supp. 3d 1373 (Ct. Int'l Tr. 2019). SeAH also argued that Commerce improperly rejected its case brief, which contained citations to certain academic texts not part of the administrative record. Id. at 1300–03; Letter from Commerce Rejecting SeAH's Sept. 1, 2015 Case Br., 1–2, PD 384, bar code 3302027-01 (Sept. 3, 2015); [SeAH's] Case Br., PD 377–79, bar codes 3301610-01–03 (Sept. 1, 2015) ("SeAH's Rejected Brief").

Consol. Court No. 15-00334

Page 4

This court sustained Commerce’s determinations with respect to its use of differential pricing analysis and rejection of SeAH’s case brief. Stupp I, 359 F. Supp. 3d at 1299–1306. Specifically, the court found that Commerce correctly rejected SeAH’s brief because the academic authorities cited in the brief constituted new factual information intended to advance SeAH’s arguments. Id. at 1301. The court also found that Commerce’s differential pricing analysis was supported by substantial evidence because, among other reasons, Commerce was not required to apply the Cohen’s *d* test in accordance with academic literature. Id. at 1302–06.

The Court of Appeals remanded, instructing Commerce to further explain why its use of the Cohen’s *d* test was reasonable in light of “significant concerns” related to application of the test. Stupp III, 5 F. 4th at 1357. Specifically, the Court of Appeals questioned the reasonableness of Commerce’s application of Cohen’s *d* test to data failing to satisfy the statistical criteria of normality, equal variance, and sufficient observation size. Id. 1357–60. Citing to academic literature examining the use of Cohen’s *d* test to measure effect size, the Court of Appeals expressed concern that Commerce’s failure to satisfy the statistical criteria assumed by Cohen’s test could “undermine the usefulness of the interpretive cutoffs,” resulting in artificially inflated dumping margins. Id. at 1357. The Court of Appeals affirmed the remaining issues from Stupp I, including this court’s decision to uphold Commerce’s rejection of SeAH’s case brief. Id. at 1344.

JURISDICTION AND STANDARD OF REVIEW

The court has jurisdiction pursuant to 28 U.S.C. § 1581(c) (2018), which grants the court authority to review actions initiated under 19 U.S.C. § 1516a(a)(2)(B)(i)¹ contesting the final determination in an antidumping duty order. The court will uphold Commerce’s determination unless it is “unsupported by substantial evidence on the record, or otherwise not in accordance with law.” 19 U.S.C. § 1516a(b)(1)(B)(i). “The results of a redetermination pursuant to court remand are also reviewed ‘for compliance with the court’s remand order.’” Xinjiamei Furniture Co. v. United States, 968 F. Supp. 2d 1255, 1259 (Ct. Int’l Tr. 2014).

DISCUSSION

On remand, SeAH challenges Commerce’s application of the Cohen’s *d* test on the grounds that (1) assumptions underlying the test have not been met, (2) the large cutoff prescribed by the test is arbitrary, and (3) random variables such as exchange rates can cause “false positives.” See Cmts. of [SeAH] on Final Determin. on Remand, 5–36, June 14, 2022, ECF No. 216 (“SeAH’s Cmts.”). Defendant and Welspun Tubular LLC (“Welspun”) counter that (1) the assumptions are inapplicable, (2) Commerce’s application of Cohen’s *d* test leads to reasonable results, (3) the cutoff is supported by statistical literature, (4) SeAH cannot introduce non-record documents for the first time on remand, and (5) SeAH failed to exhaust administrative remedies for its

¹ Further citations to the Tariff Act of 1930, as amended, are to the relevant provisions of Title 19 of the U.S. Code, 2018 edition.

Consol. Court No. 15-00334

Page 6

exchange rate-related arguments. See Def.'s Corr. Resp. to Cmts., 9–34, Sept. 22, 2022, ECF No. 230 (“Def.’s Reply”); [Welspun’s] Reply [SeAH’s] Cmts. on Remand Redeterm., 18–32, Aug. 15, 2022, ECF No. 218 (“Welspun’s Reply”). For the following reasons, the court sustains the results of Commerce’s remand redetermination.

I. SeAH’s Non-Record Documents

SeAH’s comments to the Remand Results reference several pieces of academic literature which were not included in the administrative record. See SeAH’s Cmts. at 6–36. Welspun and Defendant argue that the court should disregard these materials, as judicial review is limited to the agency record. Welspun’s Reply at 19–20; Def.’s Reply at 10–12. SeAH argues that the court may take judicial notice, or otherwise consider, these materials to better understand the statistical principles behind Cohen’s *d* test. Reply of [SeAH] to Responses by Def. and [Welspun], 10–11, Sept. 28, 2022, ECF No. 236 (“SeAH’s Reply”). For the following reasons, the court need not take judicial notice of SeAH’s non-record documents to understand the statistical principles they illustrate.

Judicial review is generally limited to the administrative record before the agency at the time it rendered its decision. See Camp v. Pitts, 411 U.S. 138, 142 (1973). “The purpose of limiting review to the record actually before the agency is to guard against courts using new evidence to ‘convert the “arbitrary and capricious” standard into effectively de novo review.’” Axiom Res. Mgmt, Inc. v. United States,

Consol. Court No. 15-00334

Page 7

564 F.3d 1374, 1380 (Fed. Cir. 2009) (quoting Murakami v. United States, 46 Fed. Cl. 731, 735 (2000), aff'd, 398 F.3d 1342 (Fed. Cir. 2005)).

This court previously upheld Commerce’s decision to reject SeAH’s non-record documents, on the grounds that the submissions constituted new factual information not on the administrative record. Stupp I, 359 F. Supp. 3d at 1299–1306. SeAH’s submissions primarily cited academic articles relating to application of the Cohen’s *d* test under certain conditions. SeAH’s Rejected Brief at 26–33. On appeal, the Court of Appeals affirmed this court’s decision rejecting the non-record information, concluding that SeAH’s materials were not introduced to correct inaccuracies in Commerce’s reporting, but to support its argument challenging Commerce’s use of Cohen’s *d* test. Stupp III, 5 F.4th at 1350. In Stupp III, the Court of Appeals nevertheless referenced and quoted from several of the non-record texts introduced by SeAH.² Id. at 1357–59. On remand, Commerce asked SeAH to place the

² The Federal Circuit cited the following five works: Grissom, Robert and Kim, John, Effect Sizes for Research: Univariate and Multivariate Applications (2nd ed. 2012), A-580-876, PRRD 8, bar code 4181776-01 (Nov. 12, 2021) (“Grissom & Kim”); Coe, Robert, It’s the Effect Size Stupid: What Effect Size Is and Why It Is Important, paper presented at the Annual Conference of the British Educational Research Association (September 2002), A-580-876, PRRD 8, bar code 4181776-01 (Nov. 12, 2021) (“Coe”); Lane, David, et al., Introduction to Statistics, Online Edition, A-580-876, PRRD 8, bar code 4181776-01 (Nov. 12, 2021) (“Lane”); Algina, James, Keselman, H.J., and Penfield, Randall, An Alternative to Cohen’s Standardized Mean Difference Effect Size: A Robust Parameter and Confidence Interval in the Two Independent Groups Case, 10 *Psychological Methods* (2005), A-580-876, PRRD 8, bar code 4181776-01 (Nov. 12, 2021) (“Algina”); Li, Johnson Ching-Hong, Effect Size Measures in a Two-Independent-Samples Case With Nonnormal and Nonhomogenous Data, *Behavioral Research* (2015), A-580-876, PRRD 8, bar code 4181776-01 (Nov. 12, 2021) (“Li”).

Consol. Court No. 15-00334

Page 8

previously-rejected materials on the record, which SeAH did. See Letter from [Commerce] to Interested Parties, A-580-876, PRRD 1, bar code 4176823-01 (Oct. 29, 2021); SeAH Submission of Publications Requested, A-580-876, PRRD 8, bar code 4181776-01 (Nov. 12, 2021). In its comments on the remand redetermination, SeAH again cites new academic sources not on the record, arguing that the court may consider the underlying statistical principles which the texts discuss.³ See SeAH's Cmts. at 6–25. SeAH claims that the Court of Appeals considered SeAH's previous academic sources in Stupp III, despite upholding Commerce's rejection of SeAH's brief which contained these materials. SeAH's Reply at 10–11. Although SeAH states that the Court of Appeals took judicial notice of the texts, it later clarified that the court may consider the statistical principles regardless of whether the texts themselves are on the record. Response of [SeAH] to Def's Sur-Reply, 2–3, Nov. 14, 2022, ECF No. 247 ("SeAH's Sur-Reply").⁴

³ SeAH cites to the following six non-record sources in its comments: Todd D. Little, Oxford Handbook of Quantitative Methods in Psychology (2013); Ricca and Blaine, Notes on a Nonparametric Estimate of Effect Size, 90:1 *Journal of Experimental Education* 249 (2022); Hedges and Olkin, Overlap Between Treatment and Control Distributions as an Effect Size Measure in Experiments, 21:1 *Psychological Methods* 61 (2016); Huberty and Lohman, Group Overlap as a Basis for Effect Size, 60:4 *Educational and Psychological Measurement* 543 (2000); J. Cohen, A Power Primer, 112:1 *Psychological Bulletin* 155 (1992); F. Alvarez, A. Atkeson, and P. Kehoe, If Exchange Rates Are Random Walks, Then Almost Everything We Say about Monetary Policy is Wrong, Federal Reserve Bank Of Minneapolis Research Department Staff Report 388 (2007).

⁴ SeAH argues that the Court of Appeals' decision "stands for the proposition that, when an agency purports to be using a statistical test in accordance with widely-adopted statistical practice, the courts may consider non-record academic materials to evaluate that claim." SeAH's Sur-Reply at 2.

Consol. Court No. 15-00334

Page 9

Consistent with the approach of the Court of Appeals, the court may recognize the basic statistical principles discussed in these texts. The idea, for example, that a skewed statistical sample may yield inaccurate results is inductive reasoning—not an assertion of fact. The Court of Appeals’ references to academia do not render its reasoning dependent on academic sources. Thus, the court considers Commerce’s Cohen’s *d* methodology in the same way it would review any other methodology, and may make logical inferences without taking judicial notice of SeAH’s literature.

II. Administrative Exhaustion

SeAH argues that random fluctuations of exchange rates can affect the Cohen’s *d* test, and lead to inaccurate results. SeAH’s Cmts. at 24–28. Welspun and Defendant argue SeAH failed to properly exhaust this argument. Welspun’s Reply at 20–21; Def.’s Reply. at 27–28. For the following reasons, the court concludes that SeAH has exhausted this argument.

Pursuant to 28 U.S.C. § 2637(d), the court “shall, where appropriate, require the exhaustion of administrative remedies,” including at the preliminary determination stage before the agency. 28 U.S.C. § 2637(d); 19 C.F.R § 351.309(c)(2). Section 2637(d) grants the court “discretion to identify circumstances where exhaustion of administrative remedies does not apply.” ABB, Inc. v. United States, 920 F.3d 811, 818 (Fed. Cir. 2019) (quoting Consol. Bearings Co. v. United States, 348 F.3d 997, 1003 (Fed. Cir. 2003)). The court may also excuse exhaustion in certain

Consol. Court No. 15-00334

Page 10

circumstances, such as when a party is raising a “pure question of law.” Agro Dutch Indus. Ltd. v. United States, 508 F.3d 1024, 1029 (Fed. Cir. 2007).

In its comments on the draft redetermination, SeAH argued that random factors, such as exchange rates, could cause the standard deviation of test populations to vary significantly. SeAH’s Cmts. on Draft Redetermination, 17–20, A-580-876, PRRD 30, bar code 4224356-02 (March 21, 2022). SeAH did not provide an exchange rate table, or assert that its actual sales during the period of review were affected by these factors. See id. Subsequently, in its comments on the final remand results, SeAH again argues that Cohen’s d could be significantly affected by random factors where the population of data is not normally distributed. SeAH’s Cmts. at 24–28. SeAH adds that its sales were, in fact, affected by fluctuations in the exchange rate between the U.S. dollar and Korean won, because its inland freight expenses were denominated in won. Id. at 24. Welspun counters that SeAH failed to raise its exchange rate argument and supporting factual information during the draft redetermination. Welspun’s Reply at 26.

SeAH has exhausted its exchange rate argument. SeAH’s exchange rate examples provide an illustration of how it believes random factors can render the Cohen’s d test inaccurate when values are not normally distributed. Normal distributions is one of the three assumptions that the Court of Appeals remanded to Commerce to explain. See Stupp III, 5 F. 4th at 1360. Therefore, Welspun’s argument that Commerce had no opportunity to address SeAH’s exchange rate calculations

Consol. Court No. 15-00334

Page 11

misses the point; these calculations are not a new argument, but an illustration of the same scenario Commerce was directed to explain.

SeAH separately argues Commerce must “ignore” fluctuations in exchange rates pursuant to 19 U.S.C. 1677b-1(a). SeAH’s Br. at 8. SeAH concedes that it did not raise this argument in its comments to the draft remand results; nevertheless, it argues that this argument may be considered as a “pure question of law.” Oral Argument at 0:27:39–0:27:53. Defendant argues that Commerce had no opportunity to consider this argument on remand, and Welspun characterizes the argument as a mixed question of law and fact. Oral Argument at 0:24:15–0:26:33, 0:27:56–0:28:39. Whether § 1677b-1(a) is pertinent to Commerce’s differential pricing analysis is a matter of statutory interpretation, not subject to exhaustion requirements. See Agro Dutch Indus., 508 F.3d at 1029. However, SeAH’s argument that § 1677b-1(a) directs Commerce to compensate for exchange rate variations is inapposite. In its full context the statute directs Commerce to use the exchange rate “in effect on the date of sale” for valuation of merchandise, and to ignore fluctuations on that particular date. 19 U.S.C. 1677b-1(a). The plain language does not mandate that Commerce compensate for a respondent’s decision to report expenses in a foreign currency, as SeAH suggests.

III. Differential Pricing Analysis

In Stupp III, the Court of Appeals remanded for further explanation of Commerce’s application of the Cohen’s *d* test as part of its differential pricing

Consol. Court No. 15-00334

Page 12

analysis. Stupp III, 5 F.4th at 1360. On remand, SeAH renews its argument that Commerce's application of Cohen's *d* test is flawed because it fails to take into account assumptions of sample size, distribution, and variance underlying the test, and further argues Commerce's choice of Cohen's large cutoff is arbitrary. SeAH's Cmts. at 6–24. SeAH also claims random fluctuations in exchange rates can affect the *d* coefficient, causing even test groups with identical prices to pass. Id. at 24–28. Commerce counters that its Cohen's *d* analysis does not operate in a vacuum, and must be considered with the ratio test and meaningful difference test. See Remand Results at 26, 28, 30–31, 41–42, 54–60. Commerce also argues the cutoffs are tied to real-world criteria, that small fluctuations in price will not lead to “false positives” in Cohen's test, and that use of the 0.8 threshold results in reasonably infrequent application of alternative methodologies. Remand Results at 16–19, 32, 54–60. For the following reasons, Commerce has adequately addressed Court of Appeals' concerns.

When investigating whether subject merchandise is being sold at less than fair value, Commerce typically compares “the weighted average of the normal values to the weighted average of the export (and constructed export prices) for comparable merchandise” unless it determines another method is appropriate. 19 U.S.C. § 1677f-1(d)(1)(A)(i); 19 C.F.R. § 351.414(c)(1). This average-to-average (“A-to-A”) method compares the weighted average of a respondent's home country sales prices during the investigation period to the weighted average of the respondent's sales prices in

Consol. Court No. 15-00334

Page 13

the United States during the same period. See 19 C.F.R. § 351.414(b)(1). One concern with the A-to-A method is that it may allow a foreign producer or exporter to engage in “targeted dumping,” which occurs when an exporter sells at a dumped price to particular customers or regions, while selling at higher prices to other customers or regions. See Apex Frozen Foods Priv. Ltd. v. United States, 862 F.3d 1337, 1341 (Fed. Cir. 2017) (“Apex II”). As a result, higher-priced products can mask dumped products when Commerce averages the sales using the A-to-A method.

Congress addressed concerns over targeted dumping with the passage of 19 U.S.C. § 1677f-1(d)(1)(B). See Apex II, 862 F.3d at 1342. Section 1677f-1(d)(1)(B) allows Commerce to compare “the weighted average of the normal values to export prices . . . of individual transactions for comparable merchandise if (i) there is a pattern of export prices . . . for comparable merchandise that differ significantly among purchasers, regions or periods of time, and (ii) [Commerce] explains why such differences cannot be taken into account using [the A-to-A method or transaction-to-transaction method⁵].” 19 U.S.C. § 1677f-1(d)(1)(B)(i)–(ii). Targeted dumping is more likely when export prices fit a pricing model that differs significantly across different market segments. Apex II, 862 F.3d at 1341–42. Congress has not provided a method for Commerce to use to determine whether a pattern of significantly different prices

⁵ Commerce’s regulations provide that the transaction-to-transaction method, which compares prices of individual transactions, will be employed only in rare cases, “such as when there are very few sales of subject merchandise and the merchandise sold in each market is identical or very similar or is custom-made.” 19 C.F.R. § 351.414(c)(2).

exists. However, the Statement of Administrative Action (“SAA”) of the Uruguay Round Agreements Act explains that Commerce should proceed “on a case-by-case basis, because small differences may be significant for one industry or one type of product, but not for another.” Uruguay Round Agreements Act, Statement of Administrative Action, H.R. Doc. No. 103-316, vol. 1, at 842–43 (1994), reprinted in 1994 U.S.C.C.A.N. 4040, 4178.⁶

To determine whether the criteria set forth in § 1677f-1(d)(1)(B) are met, Commerce conducts a “differential pricing analysis” of a respondent’s sales. See Differential Pricing Analysis; Request for Comments, 79 Fed. Reg. 26,720, 26,722 (Dep’t of Commerce May 9, 2014). This analysis contains three tests. First, Commerce applies to respondent’s sales what it refers to as the “Cohen’s *d* test,” described in more detail below, which measures the degree of price disparity between groups of sales. Id. Commerce then counts the percentage of sales by value which “pass” the Cohen’s *d* test, and applies its “ratio test.” Id. at 26,722–23. If 33% of respondent’s sales or less pass, Commerce uses the A-to-A method, and if 66% or more pass, Commerce uses the A-to-T method. Id. If the total percentage of passing sales is between 33% and 66%, Commerce takes a hybrid approach, applying the A-to-T method to those sales passing the test, and the A-to-A method to the remainder. Id. Finally, if Commerce has not selected the A-to-A method for all sales, it applies

⁶ The SAA is an “authoritative expression by the United States concerning the interpretation and application” of the Uruguay Rounds Agreement Act. 19 U.S.C. § 3512(d).

Consol. Court No. 15-00334

Page 15

the “meaningful difference” test to determine whether the A-to-A method could nevertheless account for the disparate pricing. Id. at 26,723. Commerce applies the test by comparing a respondent’s dumping margin using both A-to-A and the selected method. Id. If the A-to-A margin is below the de minimis threshold and the margin from the selected method is not, or if both margins are above the threshold and differ by 25% or more, Commerce continues to use the selected method; otherwise, Commerce applies the A-to-A method for all sales. Id.

As applied by Commerce, the Cohen’s *d* test involves comparing the product-specific prices of “test groups” of a respondent’s sales to a “comparison group” by region, purchaser, and time period. Stupp III, 5 F.4th at 1346. For each category, Commerce segregates sales into subsets, with one subset becoming the test group, and the remaining subsets being combined as the comparison group. Id. Commerce then calculates the means and standard deviations of the test and comparison groups. Id. Commerce then calculates a Cohen’s *d* coefficient by dividing the difference in the groups’ means by the groups’ standard deviation.⁷ Id. Each subset is thus tested against the remaining subsets across each category, and assigned a *d* coefficient by

⁷ Thus, $d = |\text{mean of test group} - \text{mean of control group}| \div \text{standard deviation}$. Commerce uses a modified version of this formula, substituting the square root of the simple average of the groups’ variances for standard deviation. The Cohen’s *d* test solves for a coefficient representing “effect size.” See generally Cohen, Jacob, Statistical Power Analysis for the Behavioral Sciences, (2nd ed. 1988), A-580-876, PRRD 8, bar code 4181776-01 (Nov. 12, 2021) (“Cohen”). “Effect size quantifies the size of the difference between two groups, and may therefore be said to be a true measure of the significance of the difference.” Coe at 7.

Consol. Court No. 15-00334

Page 16

solving Cohen's ratio. If the d value of a test group is equal to or greater than the "large threshold," or 0.8, the observations within that group are said to have "passed" the Cohen's d test. Id. at 1347.

In Stupp III, the Court of Appeals expressed concern that Commerce's application of Cohen's d under certain circumstances could undermine the usefulness of the test in less-than-fair-value determinations. Specifically, the Court of Appeals identified three potential scenarios in which use of Cohen's d could be problematic: first, when the distribution of a respondent's sales data is not normal, second, when the test groups have few data points, and third, when there is minimal variance in a respondent's sales. Stupp III, 5 F.4th at 1357–59. The assumption of normality is satisfied when a fixed percentage of the population falls within each standard deviation from the mean—in other words, that a population density graph generally shows a symmetrical, bell-shaped curve. See Starnes, Yates, and Moore, Statistics through Applications, 116 (2005), A-580-876, PRRD 8, bar code 4181776-01 (Nov. 12, 2021). The assumption of size is satisfied when the population is sufficiently large. See Cohen at 21. The assumption of homogeneous variances is satisfied when the standard deviations of test and comparison groups are similar. See Grissom at 68–69. Commerce argued in Stupp III, as it does now, that the three assumptions are only relevant as a matter of statistical significance, and do not apply when analyzing a whole population; the Court of Appeals concluded that this explanation did not fully address its concerns. Stupp III, 5 F.4th at 1360.

The Court of Appeals illustrated the problems it identified with the Cohen’s *d* test through two hypotheticals. First, the Court of Appeals considered a situation in which Commerce analyzed a group of only eight export sales across four groups, such that each test group would consist of only two sales. Stupp III, 5 F. 4th at 1358–59. With groups of such small numbers, the Court of Appeals pointed out that there would be some upward bias in effect size, such that the test would produce more “passing” results, and potentially exaggerate dumping margins. Id. at 1359. The Court of Appeals also observed that a group of only two sales would lack normality. Id. Second, the Court of Appeals described a test group of five sales of about \$100 each, which differed from one-another by up to two cents. Id. Because the standard deviation of such a group would be so small, the Court of Appeals pointed out that the denominator in Cohen’s ratio would be drastically reduced, again causing an increase in effect size, and inflating the resulting dumping margin. Id. The Court of Appeals noted that an objective examiner looking at these similar sales prices “would be unlikely to conclude that they embody a ‘pattern’ of prices which ‘differ significantly.’” Id. (citing 19 U.S.C. § 1677f-1(d)(1)(B)(i)).

Commerce reasonably explains that Cohen’s *d* test does not operate in a vacuum, but as part of the differential pricing analysis as a whole.⁸ Turning first to

⁸ The parties devote a significant part of their briefings to discussion of (1) the permissibility of using Cohen’s *d* test on full populations, and (2) questions of

(footnote continued)

the assumptions of population size and normalcy, the Court of Appeals questioned whether small sample sizes without normal distributions could “exaggerate” dumping margins by introducing an “upward bias” to effect size. Stupp III, 5 F.4th at 1359.⁹ Addressing the Court of Appeals’ concerns about population size, Commerce

statistical significance versus practical significance. See Remand Results at 11–16, 43–51; SeAH’s Cmts. at 7–10; Def.’s Br. at 12–20; Welspun’s Br. at 21–23; SeAH’s Reply at 17–27; Def’s Sur-Reply at 12–29, 34–35; SeAH’s Sur-Reply at 16–24. Neither question is determinative of whether Commerce’s methodology is reasonable. Both arguments have already been raised before the Court of Appeals, which concluded that they did not resolve its concern over whether the absence of certain assumptions forecloses Commerce’s use of Cohen’s *d* test. Stupp III, 5 F.4th at 1360.

Commerce correctly asserts that a “t-test” for statistical significance is used with sampled data, and that Dr. Cohen considered normal distribution and equal variance as necessary assumptions in a t-test. See Cohen at 19; Remand Results at 12–16. However, Commerce improperly reasons that because there is no need for a t-test, there is no basis for the assumptions. Remand Results at 14. Commerce also asserts that SeAH’s assumptions are only relevant as a matter of statistical significance, and that they do not apply because Cohen’s *d* test determines practical significance. Remand Results at 14, 43–45. That these assumptions are required for questions of statistical significance does not answer the question of whether they are also needed to determine practical significance, as the Court of Appeals suggests. See Stupp III, 5 F.4th at 1360.

Although SeAH claims that academic sources do not support Commerce’s use of Cohen’s *d* in its differential pricing analysis, this argument is inapposite. SeAH’s decision to substantially advance its arguments using labels taken from statistical literature does not alter the court’s obligation on review. See Soc Trang Seafood Joint Stock Co. v. United States, 321 F. Supp. 3d 1329, 1339 n.13 (2018) (“the fact that Commerce has adopted a methodology based upon a statistical tool known as Cohen’s *d*, and chooses to refer to this methodology as Cohen’s *d*, does not diminish the discretion granted to Commerce”); see also Mid Continent Steel & Wire, Inc. v. United States, 31 F.4th 1367 (Fed. Cir. 2022) (“Commerce’s job is not to follow a statistical test as explained in published literature for its own sake, but to implement the statutory mandate to determine when prices of certain groups “differ significantly”).

⁹ Although the parties dispute whether such results are really “false positives,” it is undisputed that in at least some instances, groups with as few as two sales have

(footnote continued)

Consol. Court No. 15-00334

Page 19

explains that its Cohen's d analysis does not stand alone, and operates together with the ratio test and meaningful difference test. See Remand Results at 26, 28, 30–31, 41–42, 54–60. Thus, even if the Cohen's d values of small test groups were less accurate than for large test groups, this difference does not by itself render Commerce's use of Cohen's test unreasonable, because the ratio test and meaningful difference test compensate for inaccuracies. See id. Commerce's differential pricing analysis looks at the frequency and impact of effect size to detect targeted dumping—not the effect size alone. See Cohen at 8; Remand Results at 26–28. As Commerce points out in its remand redetermination, the “sole purpose of the Cohen's d test” is to determine whether prices “differ significantly” across region, time period, or customer. Remand Results at 41. The “pattern” of export prices which Commerce must find under 19 U.S.C. § 1677f-1(d)(1)(B)(i) is then determined by the ratio test. Id. at 42. The ratio test has already been approved by the Court of Appeals, which found that Commerce's choice of the 33% and 66% thresholds was a “reasonable choice.” Stupp III, 5 F.4th at 1355. SeAH's attacks on Cohen's d test presuppose that what SeAH claims are “false positives” automatically affect the accuracy of

passed Cohen's test. See Stupp III, 5 F.4th at 1357; SeAH's Cmts. at 17–19; see Remand Results at 55, 58–59. Identifying results as “false” positives begs the question of what is a false positive. See Remand Results at 59 (“To label this result a ‘false-positive’ does not render the variances inaccurate or erroneous”). SeAH illustrates this situation using data from its own sales, showing how a group of only two sales to a single customer passed Cohen's test, despite SeAH's observation that a visual comparison of the groups on a graph showed those sales to be near the average price. SeAH's Cmts. at 18. Commerce counters that a visual inspection may be inadequate in situations involving complex calculations. Remand Results at 59.

Consol. Court No. 15-00334

Page 20

Commerce's differential pricing analysis, when in fact Commerce has allowed for 33% positives before there is any potential effect on a respondent's dumping margins.

Commerce also addresses the Court of Appeals' concern whether samples without normal distributions will produce an inappropriate number of passes. SeAH points to numerous academic sources which it claims confirm the usefulness of Cohen's test is compromised when comparing data sets with non-normal distributions. See SeAH's Cmts. at 7, n.19 (citing Cohen at 13); Id. at 12 (citing Ellis at 41); Id. at 13 (citing Starnes, Yates, and Moore at 135). The Court of Appeals has acknowledged some of these sources. See Stupp III, 5 F.4th at 1357–59 (citing Cohen at 21, Grissom & Kim at 66, Coe at 13, Lane at 645, Algina at 318, and Li at 1571). The court need not opine on the relevance of these academic observations;¹⁰ however, it logically follows that a relatively large-tailed distribution (i.e., with large standard deviation) in a test group would tend to decrease Cohen's *d* coefficient, while the opposite would result in an increase. See Remand Results at 29 (“in other words, the fat-tailed distribution may undervalue the significance of effect”) (emphasis in

¹⁰ The task of the court is not to interpret the meaning of literature treating with correct application of Cohen's *d*. Rather, the court must determine whether Commerce's methodology is reasonable in light of considerations that run counter to its decision. See Motor Vehicle Mfrs. Ass'n of U.S. v. State Farm Mut. Auto. Ins. Co., 463 U.S. 29, 43 (1983); Ceramica Regiomontana, S.A. v. United States, 636 F. Supp. 961, 966 (Ct. Int'l Tr. 1986), aff'd, 810 F.2d 1137, 1139 (Fed. Cir. 1987) (“As long as the agency's methodology and procedures are reasonable means of effectuating the statutory purpose, and there is substantial evidence in the record supporting the agency's conclusions, the court will not impose its own views as to the sufficiency of the agency's investigation or question the agency's methodology”).

Consol. Court No. 15-00334

Page 21

original). SeAH focuses on the second of these two scenarios, arguing that even inputting random data, such as exchange rates, can cause test groups to frequently pass Cohen's *d* test. SeAH's Cmts. at 25–28. SeAH further argues that the ratio test does not account for such random fluctuations.¹¹ Oral Argument at 0:42:57–0:43:43. Commerce addresses these arguments by explaining that even if Cohen's test can produce positive results under unusual circumstances, this possibility does not mean its use of Cohen's *d* is unreasonable when combined with the ratio test and meaningful difference test. See Remand Results at 26, 28, 30–31, 41–42, 54–60.

The Court of Appeals also specifically asked Commerce to explain why it can use the 0.8 threshold identified by Dr. Cohen as a measure of a significant price difference, when Commerce evaluates data which fails to meet statistical assumptions of normality, size and variance.¹² Stupp III, 5 F.4th at 1360. Although Commerce reiterates those assumptions are irrelevant, see Remand Results at 11–

¹¹ Commerce explains that, even if exchange rate fluctuations do affect prices, this effect is not “random” because a respondent can control in which currency it denominates its prices. Remand Results at 45; Oral Argument at 0:49:11–0:50:05.

¹² Although the Court of Appeals approved the 0.8 cutoff in Mid Continent Steel & Wire, Inc. v. United States, it explained in Stupp III that it had yet to consider the reasonableness of the 0.8 cutoff value when the assumptions in question have not been met. Stupp III, 5 F.4th at 1356–57 (“We held that . . . it is reasonable to adopt that [0.8] measure where there is no better objective measure of effect size. We did not, however, address SeAH’s second argument [on assumptions] in Mid Continent”) (citation omitted) discussing Mid Continent Steel & Wire, Inc. v. United States, 940 F.3d 662, 673 (Fed. Cir. 2019); see also Oral Argument at 1:39:45–1:40:30. More specifically, SeAH argues that it is unreasonable to compare its prices, which are not normally distributed, using a subjective benchmark that was derived from a normally-distributed population. SeAH’s Cmts. at 10–12.

Consol. Court No. 15-00334

Page 22

16, it also explains its choice of the 0.8 threshold as a function of its differential pricing analysis. First, Commerce explains that it employs the 0.8 threshold to identify where prices “differ significantly” pursuant to 19 U.S.C. § 1677f-1(d)(1)(B)(i). Remand Results at 11. Second, Commerce states the 0.8 measurement “represents a difference which is ‘grossly perceptible.’” Remand Results at 52. Reasonably discernible from this statement is that Commerce considers a significant difference to be grossly perceptible in the same way that Dr. Cohen identified a large threshold as one that is “grossly perceptible.” See Cohen at 27. The SAA to the Uruguay Round Agreements Act directs Commerce to proceed “on a case-by-case basis, because small differences may be significant for one industry or one type of product, but not for another,” SAA at 842–43; thus, Commerce’s choice of a measurement that is a function of standard deviation as a uniform approach to identify differences as significant is reasonable, even if the absolute difference in means is small. Commerce’s approach tailors the question of what is a significant difference in price to the pricing parameters of different products. Third, Commerce adequately explains its adoption of Cohen’s widely-recognized choice of 0.8 as a large threshold as significant. Remand Results at 18. It explains that it chose the 0.8 standard because it was “a conservative standard to determine that the observed price differences are significant.” Id. Commerce summarizes its reasoning by explaining that “[u]sing Dr. Cohen’s thresholds is a reasonable approach to interpret whether the difference in the prices is significant and the further interpretation of the

Consol. Court No. 15-00334

Page 23

difference in the prices in the context of the calculation of dumping margins ensures the reasonable and limited application of the alternative comparison methodology.” Id. at 33. Thus, Commerce chose a threshold it predicted would result in limited application of the alternative methodology.

Although Commerce adopted this yardstick from Dr. Cohen, and did so because it was widely acknowledged in the statistical literature, Commerce does not rely on the prominence of this yardstick alone. Commerce elaborates that its “actual application of the Cohen’s *d* test in the context of the differential pricing analysis resulted in the application of an alternative comparison methodology to a relatively small number of respondents.” Remand Results at 32. Discernible from Commerce’s explanation is that the 0.8 cutoff produces reasonable passing rates once the ratio and meaningful difference tests are applied. SeAH challenges Commerce’s reliance on the 0.8 threshold as large, arguing that Commerce’s only basis for using the threshold is that it is widely accepted. SeAH’s Cmts. at 10–11. However, in addition to relying on a widely-accepted standard for “grossly perceptible” to determine what is significant, Commerce defines “significant” with reference to the impact a price difference has on a respondent’s dumping margins. Remand Results at 32. Finding that the 0.8 threshold leads to relatively few determinations of targeted dumping, Commerce concludes that its choice is reasonable. Id.

Congress delegated to Commerce the authority to determine where a price difference is significant. 19 U.S.C. § 1677f-1(d)(1)(B)(i). Congress also made clear

Consol. Court No. 15-00334

Page 24

that the definition of a “significant price difference” would depend on the product at issue. See SAA at 842–43. Thus, Congress entrusted Commerce to use its expertise and knowledge of pricing to gauge price distinctions. Cf. Fujitsu General Ltd. v. United States, 88 F.3d 1034, 1039 (Fed. Cir. 1996) (granting Commerce significant deference in determinations “involv[ing] complex economic and accounting decisions of a technical nature”). Commerce’s decision to adopt Cohen’s 0.8 (“large”) threshold as a measure of significance because it is widely accepted in the statistical literature does not undermine the reasonableness of that choice, if it is based on Commerce’s expertise and Commerce demonstrates the reasonableness of that choice with reference to the impact it has on the differential pricing analysis. Thus, Commerce’s reference to Cohen’s work does not circumscribe its discretion to choose the same values in a new context, because that choice is itself reasonable.

Commerce addresses the Court of Appeals’ concern that prices with small variances, which hover around the same value, will produce inaccurate results on Cohen’s test. As an initial matter, Commerce explains that results which pass Cohen’s test under these circumstances are not “false positives,” as small differences in average prices will mean that variances, too, will be small. Remand Results at 59; Oral Argument at 1:06:32–1:07:02. Thus, it is discernable that a small variance means a small difference in price will be more significant, and a passing result under these circumstances is not necessarily “erroneous.” Remand Results at 59. Nevertheless, the Court of Appeals observed that an objective examiner considering

Consol. Court No. 15-00334

Page 25

a group of sales where prices differed by only a few cents would be unlikely to conclude that they show a “pattern” of prices that “differ significantly” under the statute. Stupp III, 5 F.4th at 1359. Commerce responds to this issue by pointing out that an examiner would indeed conclude that there was no pattern—because Commerce does not look for a pattern at this stage of its differential pricing analysis. Remand Results at 41. Again, Commerce explains that the ratio test determines whether a pattern exists, while Cohen’s *d* test only shows whether there are significant price differences. Id. at 41–42. Thus, Cohen’s test would need to generate enough “false positives” to overcome the 33% threshold, at minimum, and there is no evidence on the record suggesting that price patterns, such as that proposed by the Court of Appeals, occur with frequency in SeAH’s sales.

Additionally, to specifically address the hypothetical proposed by the Court of Appeals, Commerce explains that, in addition to the ratio test, the meaningful difference test would prevent low-variance sales which pass Cohen’s *d* test from impacting a respondent’s dumping margins. See Remand Results at 30–31. Adopting the Court of Appeals’ example in which all of a respondent’s prices hovered around \$100 and passed Cohen’s test, Commerce explains that even in this extreme scenario, the respondent would still be assessed under the A-to-A method. Id. Choosing a normal value for comparison equal to the highest sales price, and thus maximizing the respondent’s theoretical dumping margin, Commerce observes the margin would still be under the 2% de minimis threshold. Id.; see 19 U.S.C § 1673d(a)(4),

1673b(b)(3). SeAH argues that Commerce’s reliance on the meaningful difference test is misplaced, because even changes of less than 2% in a respondent’s dumping margin can cross the de minimis threshold and result in a “meaningful difference” finding. SeAH’s Sur-Reply at 15. Specifically, SeAH argues that when the Cohen’s *d* results from small-variance data sets of different products are cumulated, Commerce may find that a respondent’s sales pass the thresholds for both ratio test and the meaningful difference test, even if price differences are negligible. Oral Argument at 1:13:37–1:16:06.¹³ This argument overstates Commerce’s position. It is reasonably discernable that Commerce does not rely on the meaningful difference test to prevent all “inappropriate” passes from affecting a respondent’s dumping margins. Commerce has explained the meaningful difference test compensates for a specific concern with low-variance sales which the Court of Appeals identified. See Stupp III, 5 F.4th at 1359; Remand Results at 30–31. Moreover, SeAH’s argument is misplaced, because the question before the court is not whether it is possible to construct an unusual scenario where Cohen’s *d* test can result in an alternative comparison method. Rather, the question is whether Commerce’s use of Cohen’s test, when applied as a component of its differential pricing analysis, is reasonable. See Ceramica Regiomontana, S.A. v. United States, 636 F. Supp. 961, 966 (Ct. Int’l Tr.

¹³ SeAH does not argue that it received an alternative method because its own combined sales inappropriately passed the Cohen’s *d* test. Rather it offers a hypothetical to challenge the reasonableness of Commerce’s methodology more generally. SeAH’s Cmts. at 8, 14–21.

Consol. Court No. 15-00334

Page 27

1986). Thus, for the forgoing reasons, Commerce has adequately explained how its methodology is reasonable.

CONCLUSION

For the foregoing reasons, Commerce's remand results are supported by substantial evidence and comply with the court's Order, Oct. 8, 2021, ECF No. 192, in conformity with the Court of Appeals' Mandate, Oct. 8, 2021, ECF No. 191, and are therefore sustained. Judgment will enter accordingly.

/s/ Claire R. Kelly
Claire R. Kelly, Judge

Dated: February 24, 2023
New York, New York

FORM 19. Certificate of Compliance with Type-Volume Limitations

Form 19
July 2020

**UNITED STATES COURT OF APPEALS
FOR THE FEDERAL CIRCUIT**

CERTIFICATE OF COMPLIANCE WITH TYPE-VOLUME LIMITATIONS

Case Number: 2023-1663

Short Case Caption: Stupp Corporation v. U.S.

Instructions: When computing a word, line, or page count, you may exclude any items listed as exempted under Fed. R. App. P. 5(c), Fed. R. App. P. 21(d), Fed. R. App. P. 27(d)(2), Fed. R. App. P. 32(f), or Fed. Cir. R. 32(b)(2).

The foregoing filing complies with the relevant type-volume limitation of the Federal Rules of Appellate Procedure and Federal Circuit Rules because it meets one of the following:

- ☒ the filing has been prepared using a proportionally-spaced typeface and includes 11,514 words.
- ☐ the filing has been prepared using a monospaced typeface and includes _____ lines of text.
- ☐ the filing contains _____ pages / _____ words / _____ lines of text, which does not exceed the maximum authorized by this court's order (ECF No. _____).

Date: 07/25/2023

Signature: /s/ Jeffrey M. Winton

Name: Jeffrey M. Winton

FORM 31. Certificate of Confidential Material

Form 31
July 2020

**UNITED STATES COURT OF APPEALS
FOR THE FEDERAL CIRCUIT**

CERTIFICATE OF CONFIDENTIAL MATERIAL

Case Number: 2023-1663

Short Case Caption: Stupp Corporation v. U.S.

Instructions: When computing a confidential word count, Fed. Cir. R. 25.1(d)(1)(C) applies the following exclusions:

- Only count each unique word or number once (repeated uses of the same word do not count more than once).
- For a responsive filing, do not count words marked confidential for the first time in the preceding filing.

The limitations of Fed. Cir. R. 25.1(d)(1) do not apply to appendices; attachments; exhibits; and addenda. *See* Fed. Cir. R. 25.1(d)(1)(D).

The foregoing document contains 24 number of unique words (including numbers) marked confidential.

- ☐ This number does not exceed the maximum of 15 words permitted by Fed. Cir. R. 25.1(d)(1)(A).
- ☒ This number does not exceed the maximum of 50 words permitted by Fed. Cir. R. 25.1(d)(1)(B) for cases under 19 U.S.C. § 1516a or 28 U.S.C. § 1491(b).
- ☐ This number exceeds the maximum permitted by Federal Circuit Rule 25.1(d)(1), and the filing is accompanied by a motion to waive the confidentiality requirements.

Date: 07/25/2023

Signature: /s/ Jeffrey M. Winton

Name: Jeffrey M. Winton